Characterisation of Naïve and Antigen-Experienced Human Antibody Repertoires

Townsend, Catherine Louise

Awarding institution:
King's College London

Characterisation of Naïve and Antigen-Experienced Human Antibody Repertoires

Catherine Louise Townsend

A Thesis Submitted to King's College London for the Degree of Doctor of Philosophy

School of Immunology and Microbial Sciences
Faculty of Life Sciences and Medicine
King's College London

October 2014 – September 2018

Unless otherwise stated, all of the work in this thesis was undertaken by the author. Where work was contributed by other people, this is indicated by footnotes.

Abstract

Antibodies have enormous diversity and form a crucial component of the humoral immune response. The relationship between antibody variable region sequences and specificity is still poorly understood, however recent advances in high-throughput sequencing technology has made the study of B cell repertoires possible. In this thesis we use, among other methods, high-throughput sequencing of B cell repertoires to examine variable region properties and specificity in a number of contexts.

First, we attempted to identify common characteristics of autoreactive/polyreactive antibody variable regions by analysing a high-throughput sequencing dataset constructed from B cell variable region sequences isolated from pre- and post-tolerance B cell repertoires. We tested our predictions using immunocytochemistry and ELISA. We found that some antibodies which show a polyreactive phenotype in ELISA do not show polyreactivity in other settings, leading us to question the definition of these antibodies and their role in the immune response.

Next, we investigated the physicochemical and structural properties of kappa and lambda light chains and the heavy chains with which they are paired. We found that the properties of the two light chain classes differ considerably, suggesting different roles in the immune response. However, we found no distinction between heavy chains paired with kappa and lambda light chains, providing no evidence of bias in heavy-light chain pairing.

Finally, we used high-throughput sequencing and ribosome display to investigate the B cell repertoires of recent Ebola survivors to determine how this disease affects the repertoire and to attempt to identify new, cross-reactive, anti-Ebola glycoprotein antibodies. The most significant repertoire feature of the Ebola survivors was the high prevalence of clonal expansions incorporating multiple heavy chain classes. Several new anti-Ebola glycoprotein single chains were isolated using ribosome display and some of these showed evidence of binding N-linked glycans on various viral glycoproteins.

Acknowledgements

Throughout the four years I have worked on this PhD, I have been fortunate enough to spend time in several laboratories both in academia and in industry. The work in this thesis would not have been possible without the help, advice and encouragement of many people at King's College London, MedImmune, The University of Surrey and Queen Mary University of London.

Firstly, I would like to thank my primary supervisor, Deborah Dunn-Walters, for all of your help and guidance over the last four years. You could always see a way forward when times seemed tough and your relentless enthusiasm for science was inspiring; particularly your ability to recite immunoglobulin genes from memory!

A big thank you also to my second supervisor, Franca Fraternali, for your expert guidance on bioinformatics and molecular modelling. I very much enjoyed learning some computational methods and I just wish I had more time to get better at it! Thank you also for making meetings so pleasant with the provision of good coffee and Italian biscuits.

Thank you to my industrial supervisor, Claire Dobson, for always making time to meet with me when I was working at MedImmune, and for being so encouraging and full of interesting ideas. You showed me that working in industry is just as exciting and fulfilling as working in academia.

I would particularly like to thank Katie Doores, my adopted fourth supervisor, for going well above and beyond what was required of you by welcoming me into your lab halfway through my PhD. It has been fascinating to learn about HIV and glycosylation and your insight on my project has been invaluable. It has been an absolute pleasure to work with you and your group over the last two years.

Many thanks also the other members of my thesis committee: Patricia Barral, Claudia Mauri and Maria Hernandes-Fuentes for your advice and guidance.

I would like to extend my gratitude to the Biotechnology and Biological Sciences Research Council (BBSRC) and to MedImmune for the BBSRC-CASE studentship which funded my project and my stipend. I would also like to acknowledge the Medical Research Council (MRC) who funded the Ebola project with which I was also involved. I would also like to thank all of the people who generously donated the blood and bone marrow which was used in this work.

To everyone involved in the MABRA consortium; I feel very fortunate as a biologist to have had the opportunity to meet regularly with physicists, mathematicians and molecular modellers and to have gained some insight into your world. I would also like to thank everyone involved

in the Ebola project at King's College London for sharing expertise and materials; it has been a privilege to have worked with you on such an important subject area.

A huge thank you to David Kipling for your bioinformatic work processing our high-throughput sequencing data, and also to Angray Kang for very generously letting me spend several weeks working in your lab at QMUL and teaching me how to conduct ribosome display.

I would like to thank all the members of the Dunn-Walters group, both at King's College London and the University of Surrey: Joselli Silva O'Hare, Victoria Martin, Tihomir Dodev, Audrey Page, Alexander Stewart and Emma Sinclair. I learned a lot from you all and I couldn't have asked for a nicer group.

I would also like to thank all the members of the Doores group at King's College London: Isabella Hüttner, Stefanie Krumm, Elda Iljazi, Luke Granger, Nelson Cano Eusebio and Karen Coss. Thank you all for being such wonderful labmates and for making me feel so welcome.

I was very fortunate to have spent several months working on my project at MedImmune, Cambridge. The time I spent there was certainly one of the highlights of my PhD; everybody was so friendly and passionate about science and always had time to debate aspects of my project. I would like to thank everyone in the department of Antibody Discovery and Protein Engineering for their help with my work and also the IgG Team for expressing some of my antibodies. I would like to express particular gratitude to James Button who is one of the best teachers and most enthusiastic scientists I have had the pleasure of working with.

Thank you to the whole Fraternali group at King's College London; I always enjoyed our occasional group outings and get-togethers. Thank you in particular to Julie Laffy for your help with antibody modelling.

I would also like to thank the wonderful colleagues I have shared an office with over the past four years, namely members of the Dunn-Walters, Spencer and Klavinskis groups. Thank you for making work such a happy place through your friendship and the near-constant supply of snacks from everyone's travels!

Finally, I would like to thank my parents and my sister for their unconditional love, support and encouragement. I would not have been able to do this without you.

Table of Contents

List of Figures

List of Tables

Abbreviations

A

AA	Amino acid(s)
ADCC	Antibody-dependent cellular cytotoxicity
ADCP	Antibody-dependent cellular phagocytosis
AID	Activation-induced cytidine deaminase
ANA	Anti-nuclear antibody
ANOVA	Analysis of Variance
APC	Antigen-presenting cell
ApoH	Apolipoprotein H
APS	Ammonium persulfate

B

BCR	B cell receptor
BDBV	Bundibugyo virus
BM	Bone Marrow
BMMC	Bone marrow mononuclear cell
BSA	Bovine serum albumin

C

CDC	Complement-dependent cytotoxicity
CDR	Complementarity determining region
cfu	Colony forming units
CH	Heavy chain constant domain
CHO	Chinese hamster ovary (cells)
Chr	Chromosome
CI	Confidence interval
CL	Light chain constant domain
CSR	Class switch recombination

D

DAPI	4',6-diamidino-2-phenylindole
dH_2O	Distilled water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol

E

EBOV	Ebola virus (Zaire)
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Endo H	Endoglycosidase H

F

Fab	Fragment antigen-binding
FACS	Fluorescence-activated cell sorting
Fc	Fragment crystallizable
FCS	Foetal calf serum
FcαR	Fc-alpha receptor
FcγR	Fc-gamma receptor
FcεR	Fc-epsilon receptor
FDC	Follicular dendritic cell
FDR	False discovery rate
FR	Framework region

G

GIF	Gastric intrinsic factor
GP	Glycoprotein
GRAVY	Grand average of hydropathicity

H

HEK 293	Human Embryonic Kidney 293
HEp-2	Human epithelial type 2
HI-FCS	Heat-inactivated foetal calf serum
His-tag	Polyhistidine-tag
HMGB1	High Mobility Group Protein 1
HP-SEC	High-pressure size exclusion chromatography
HRP	Horseradish peroxidase
HTS	High-throughput sequencing

I

IgH	Immunoglobulin heavy chain
IGHD	Immunoglobulin heavy chain diversity (gene)
IGHJ	Immunoglobulin heavy chain joining (gene)
IGHV	Immunoglobulin heavy chain variable (gene)
IgK	Immunoglobulin kappa light chain
IGKJ	Immunoglobulin kappa light chain joining (gene)
IGKV	Immunoglobulin kappa light chain variable (gene)
IgL	Immunoglobulin lambda light chain
IGLJ	Immunoglobulin lambda light chain joining (gene)
IGLV	Immunoglobulin lambda light chain variable (gene)
IMGT	International Immunogenetics Information System

K

KS (test)	Kolmogorov-Smirnov (test)

L

LB	Luria-Bertani (broth/agar)
LLOV	Lloviu virus
LPS	Lipopolysaccharide

M

mAb	Monoclonal antibody
MARV	Marburg virus
MHC	Major histocompatibility complex
MID	Multiplex identifier
MPO	Myeloperoxidase
MS	Multiple sclerosis

N

NHEJ	Non-homologous end joining
Nluc	NanoLuc
nt	Nucleotide

O

OD	Optical density
OPD	o-phenylenediamine dihydrochloride
ORF	Open reading frame

P

PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEI MAX	Polyethylenimine "Max"
PFA	Paraformaldehyde
pfu	Plaque forming units
PHA-L	Phaseolus Vulgaris Leucoagglutinin (lectin)
pI	Isoelectric point
PID	PacBio identifier
PIPE	Polymerase Incomplete Primer Extension
PNGase F	Peptide:N-glycosidase F

R

RA	Rheumatoid arthritis
RAG	Recombination-activating gene
RAVV	Marburg virus (Ravn)
RESTV	Reston virus
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute (media)
RSS	Recombination signal sequence
RT-PCR	Reverse transcription polymerase chain reaction

S

scFv	Single-chain variable fragment
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEBOV	Sudan ebolavirus
sGP	Soluble glycoprotein
SHM	Somatic hypermutation
SLE	Systemic lupus erythematosus
SlyRT	Sort Lysis Reverse Transcription
SMRT	Single-molecule real-time
SOC	Super Optimal broth with Catabolite repression (media)
ssDNA	Single stranded DNA
ssGP	Small soluble glycoprotein
SUDV	Sudan virus

T

Ta	Annealing temperature
TAE	Tris base/Acetic acid/EDTA (buffer)
TAFV	Taï Forest virus
TBE	Tris base/Boric acid/EDTA (buffer)
TCR	T cell receptor
TCS	Tissue culture supernatant
TdT	Terminal deoxynucleotidyl transferase
TEMED	Tetramethylethylenediamine
TFH	Follicular helper T cells
TI	Thymus-independent
TLR	Toll-like receptor
Tm	Melting temperature
TMB	3,3', 5,5'-Tetramethylbenzidine
TMR	Tetramethylrhodamine
TNF	Tumor necrosis factor
TY	Tryptone yeast extract (broth/agar)

U

UDG	Uracil-DNA Glycosylase
UID	Unique (mRNA) identifier

V

VZV	Varicella zoster virus

Y

YF	Yellow Fever

Chapter 1

Introduction

Chapter 1: Introduction

1.1 Antibody Structure and Function

Antibodies (also known as immunoglobulins) are Y-shaped glycoproteins produced by B cells as part of the adaptive immune response. Their main function is to bind to harmful agents such as pathogens, toxins or diseased cells and neutralise them or tag them for clearance by the immune system. Antibodies bind specifically to their targets (antigens) via variable regions at the apices of the two fragment antigen-binding (Fab) regions. They mediate effector functions via the highly conserved fragment crystallisable (Fc) region (see Figure 1-1 for structure).

In humans, the Fab regions are composed of a heavy chain (IgH) and a light chain. The variable region is composed of a variable domain from the heavy chain (VH) and a variable domain from the light chain (VL). These are held in place by the first constant domain of the heavy chain (CH1) and the constant domain of the light chain (CL). The light chain can be one of two classes: kappa (IgK) or lambda (IgL).

The Fc region is composed of the remaining constant domains of the heavy chain. Depending on the class of heavy chain (IgM, IgD, IgG, IgA, IgE) the Fc has either two (CH2, CH3) or three (CH2, CH3, CH4) domains. Each heavy chain class has a different Fc structure and so mediates different effector functions. Antibodies can either be secreted or bound to the B cell surface.

The focus of this thesis is exclusively on human antibodies unless stated otherwise.

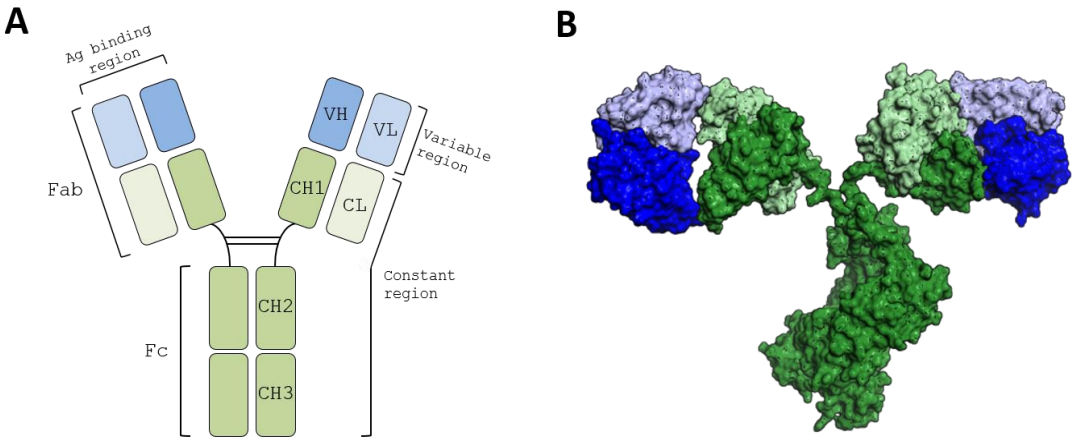


Figure 1-1: Structure of IgG antibody. (**A**) Antibodies are Y-shaped proteins composed of heavy and light chains. Antigen is bound by the extremely diverse variable region and the conserved constant regions provide structure and mediate effector functions. (**B**) Protein surface representation of an IgG antibody (PDB ID: 1IGT). In both diagrams, the variable regions are coloured blue and the constant regions are coloured green. The light chain is a lighter shade of each colour than the heavy chain.

1.2 V(D)J Recombination

The diversity of the antibody variable region is extremely high. This ensures a specific response can be mounted against any of the pathogens a host may encounter. It is estimated that the number of possible variable region sequences in the antigen-naïve B cell repertoire is greater than 10^{18}, which is more than the number of B cells in the human body (Elhanati et al., 2015). Such variable region diversity cannot be encoded solely by the germline. Instead, multiple variable region gene segments are present in the germline and these are brought together at random in developing B cells to form functional variable region sequences (B cell development is described in detail in section 1.4 page 37).

The variable region gene segments are split into three groups: Variable (V), Diversity (D) and Joining (J) (see Table 1-1 for the number of functional human VDJ genes in the germline). The process of rearranging the IgH, IgK and IgL loci to bring together VDJ genes to form functional variable region sequences is known as V(D)J recombination (see Figure 1-2). A functional heavy chain variable region is composed of a V gene, a D gene and a J gene and a functional light chain variable region is composed of a V gene and a J gene.

Table 1-1: Number of functional VDJ and constant (C) germline gene segments at the IgH, IgK and IgL loci. The chromosomal localisation (Chr) of each set of genes is shown. Table from (Lefranc & Lefranc, 2001, Lefranc, 2016).

Locus	Chr	V	D	J	C
IgH	14q32.33	38-46	23	6	9
IgK	2p11.2	34-38	0	5	1
IgL	22q11.2	29-33	0	4-5	4-5

The IgH locus is rearranged first and begins with the recombination of a D gene segment and a J gene segment; the DNA between the recombined genes is removed. The D-J pair is then recombined with an upstream V gene segment to form a primary transcript (see Figure 1-2). The IgM constant region is spliced on to the variable region to form the mRNA transcript.

The light chain loci are rearranged after the IgH locus. A single B cell may go through multiple light chain rearrangements due to the process of receptor editing (explained further in section 1.4, page 37). IgK rearrangement usually precedes IgL rearrangement (Hieter et al., 1981) and a B cell will only successfully rearrange and express either kappa light chains or lambda light chains; only very rarely does a B cell express both. Light chain loci do not have D gene segments so recombination happens between a V gene segment and a J gene segment (see Figure 1-2). The J genes and constant region genes are arranged differently at the kappa and lambda loci. At the kappa locus, there is a single constant region gene downstream of multiple

J genes. In kappa transcripts the same constant region gene is always used. At the lambda locus, each J gene has its own constant region slightly downstream (see Figure 1-2); in lambda transcripts the constant region used depends on the J gene used.

Kappa light chain usage is more common than lambda in the human peripheral blood; the proportion of antibodies using kappa light chains is reported to be between 3/5 and 2/3 (Mole et al., 1994, Brauninger et al., 2001, Montano & Morrison, 2002). However, this can differ depending on the class of antibody (Chui et al., 1990), for example, light chain usage in secreted IgA antibodies is predominantly lambda (Mole et al., 1994).

Recombination can occur multiple times if the initial rearrangement produced a non-functional or autoreactive BCR (see section 1.4, page 37, for more details). However, as the DNA between recombined gene segments is lost, further recombination can only occur using upstream V genes and downstream J genes.

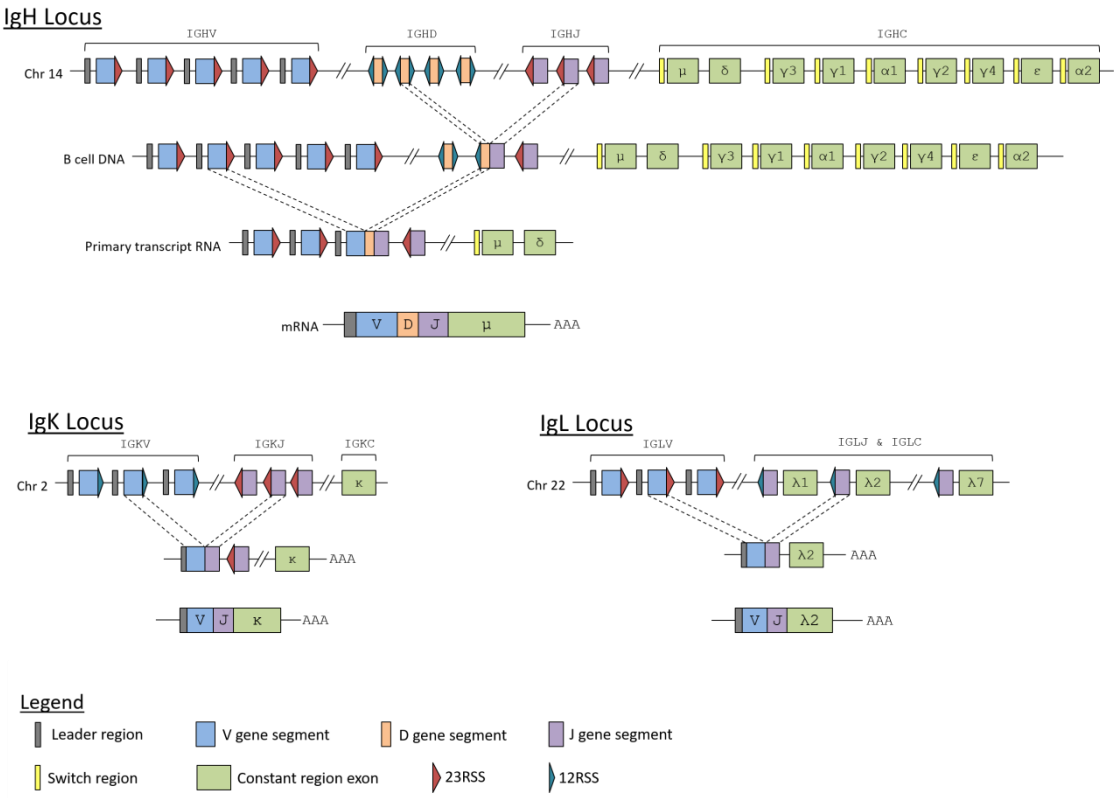


Figure 1-2: V(D)J recombination at the IgH, IgK and IgL loci. In the germline, V, D and J gene segments are arranged linearly in a non-functional state. During V(D)J recombination gene segments are brought together in a stepwise fashion. At the IgH locus, D and J genes are rearranged first, then the D-J combination is rearranged with a V gene. At the light chain loci V and J genes are rearranged. The gene segments are rearranged in the correct order using complementary recombination signal sequences (RSS). This figure is published in Dunn-Walters et al. (2018).

V(D)J recombination is controlled in part by recombination signal sequences (RSS). Each V gene, D gene and J gene is flanked by either a 12RSS or a 23RSS (see Figure 1-2). 12RSS and 23RSS are nucleotide sequences which consist of a highly conserved heptamer (5' cacagtg 3') and nonamer (5' acaaaaacc 3') separated by 12 nt and 23 nt spacer regions respectively. The heptamer end of each RSS is always adjacent to the gene. The 12-23 rule ensures that VDJ genes are brought together in the correct order; there is a strong preference for recombination between a 12RSS and a 23RSS as they are much more likely to result in open reading frames (ORFs) than recombination between two RSS of the same type (Tonegawa, 1983).

The process of recombination using RSS is mediated by the recombination-activating gene (RAG) enzyme complex; a lymphoid-specific enzyme complex composed primarily of RAG1, its cofactor RAG2 and high mobility group box 1 (HMGB1) proteins. The RAG complex binds the nonamer of one RSS and then binds a second RSS with a complementary spacer sequence (12-23 rule) (Figure 1-3A,B&C). At each RSS, the RAG complex makes a single-stranded DNA nick between the gene segment and the adjacent RSS heptamer. This results in a 3' hydroxyl group which attacks the complementary DNA strand creating a double-stranded break and covalently-sealed "hairpins" are formed at the end of each strand of DNA (Figure 1-3D) (Schatz & Ji, 2011). These double-stranded DNA breaks are then repaired by non-homologous end joining. First, the Artemis complex cleaves the "hairpin" structure to create palindromic ssDNA overhangs (referred to as P-nucleotides). Terminal deoxynucleotidyl transferase (TdT) then adds nucleotides (referred to as N-nucleotides) to each P-nucleotide ssDNA overhang until the randomly added N-nucleotides form complementary sequences. The Artemis complex removes any additional unpaired nucleotides and the gaps are joined by DNA polymerase and DNA ligase (Motea & Berdis, 2010) (Figure 1-3E). The waste DNA is released as a DNA excision circle.

The diversity created by V(D)J recombination is referred to combinatorial diversity, whereas the diversity introduced by P-nucleotides and N-nucleotides is known as junctional diversity. Further diversity can be introduced by the use of alternative IGHD reading frames and the use of two IGHD genes (Kalinina et al., 2011). Although extremely rare, there are reports of additional variation being introduced to immunoglobulin sequences by functional antibody variable regions incorporating additional pieces of DNA in a site of N nucleotide addition via interchromosomal DNA transposition from elsewhere in the genome (Tan et al., 2016).

Polymorphisms in the highly conserved RSS heptamer/nonamer sequences (Akamatsu et al., 1994) and mutations in the 12 bp or 23 bp spacer influences recombination frequency in

humans (Feeney et al., 2000, Nadel et al., 1998, Montalbano et al., 2003). This ultimately affects gene frequency in the human pre-B repertoire (Nadel et al., 1998). It is thought that greater kappa RSS similarity to a consensus sequence than lambda may explain the large kappa bias in the murine immunoglobulin repertoire (Ramsden & Wu, 1991).

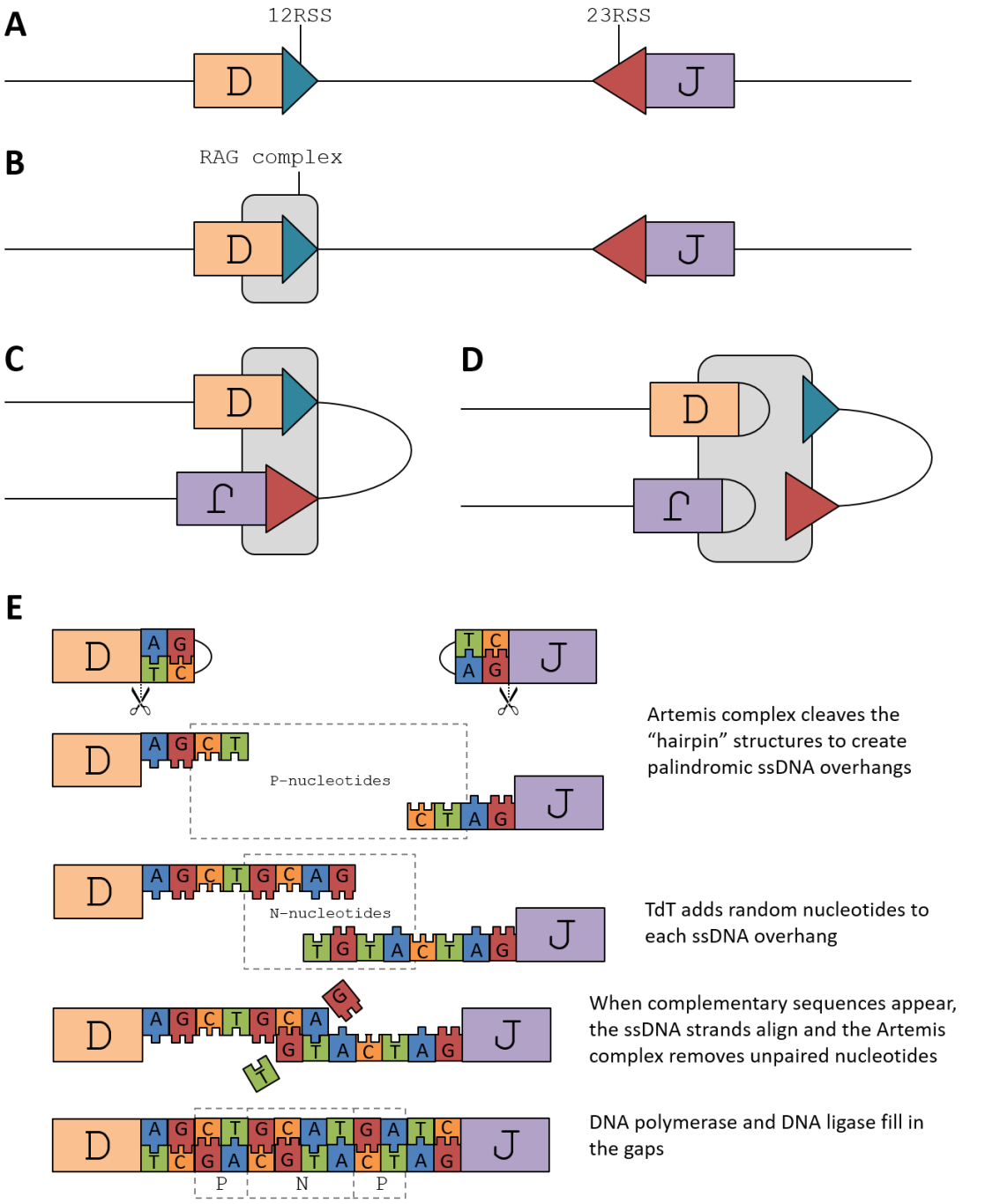


Figure 1-3: Recombination of genes by RAG recombinase. (**A**) In the genome, genes are encoded in a non-functional state. (**B**) RAG complex binds an RSS. (**C**) A complementary RSS is captured by the RAG complex. (**D**) Double-stranded breaks are created between the gene and the RSS and hairpin structures form at the ends of the nicked DNA. (**E**) Non-homologous end joining DNA repair factors join the DNA ends and TdT adds non-templated N nucleotides at the joint, increasing diversity. The wide end of each RSS arrow represents the heptamer end and the point represents the nonamer end. Figure adapted from Schatz &Ji (2011) and Motea &Berdis (2010).

1.2.1 Human Immunoglobulin Gene Nomenclature

The nomenclature used to identify VDJ genes and constant regions classifies these genes into group, subgroup, gene and allele based on their level of similarity (Lefranc, 2014) (Figure 1-4A).

Groups are genes which belong to the same multigene family within or between species. Higher vertebrates have 10 groups: IGHV, IGHD, IGHJ, IGHC, IGKV, IGKJ, IGKC, IGLV, IGLJ and IGLC. Subgroups are defined as genes in the same group which, within the species, share at least 75% identity at the nucleotide level. Groups are inter-species, whereas subgroups and genes are associated with individual species. Not all genes are functional, some are pseudogenes. Each V and D gene is identified by the subgroup to which it belongs and another number indicating which gene it is within that subgroup (e.g. IGHV3-15 and IGHV3-23 are different genes within the IGHV3 subgroup). Genes are further categorised into alleles, which are polymorphic variants of the genes. Each gene has a gene allele reference sequence (allele *01) and it is to this sequence that allelic variants are compared. Allelic variants of a gene are numbered *02, *03 etc. (Figure 1-4B).

Heavy chain constant region genes are named according to their isotype (e.g. IGHA refers to IgA constant regions, IGHG refers to the IgG constant regions etc.). Light chain kappa and lambda constant regions are identified as IGKC and IGLC respectively.

Despite the above nomenclature, in common usage (and in this thesis), constant region groups, genes and alleles are referred to as “class”, “subclass” and “allotype” respectively and VDJ subgroups are referred to as “families”. Unlike most genes, immunoglobulin genes do not need to be italicised when written (Lefranc, 2014).

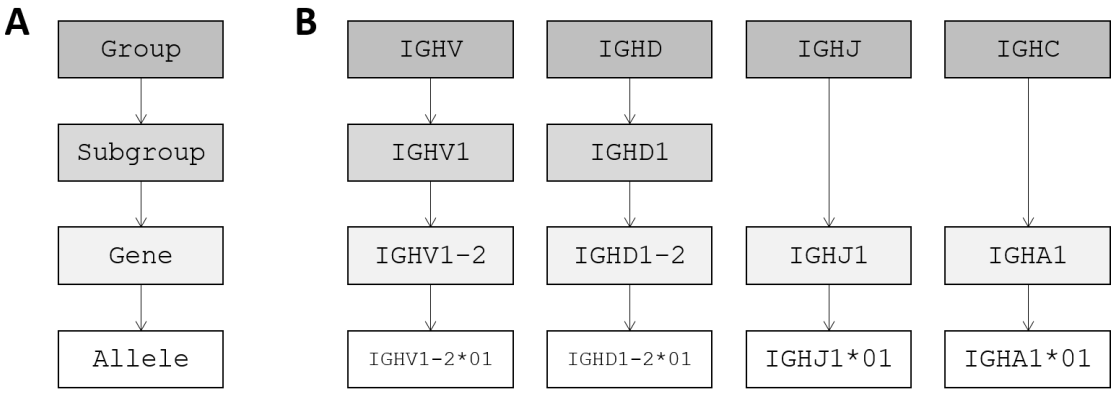


Figure 1-4: IMGT immunoglobulin gene and allele classification nomenclature. (**A**) IMGT classifies immunoglobulin genes by group, subgroup (family), gene and allele. (**B**) Example heavy chain nomenclature is shown. IGHJ and IGHC groups do not contain subgroups. Kappa and lambda light chain J and C groups also do not contain subgroups.

1.3 Complementarity-Determining Regions and Framework Regions

The successfully rearranged, functional variable region of an antibody is composed of framework regions (FR) interspersed by complementarity-determining regions (CDR). FRs are highly conserved and primarily serve a structural purpose. The CDRs are much more diverse and form loops in the variable region which make the most contact with antigen. The CDR loops are therefore most important regions for antibody specificity. The heavy chain and the light chain each contribute three CDR loops to the antigen binding site. The heavy chain CDR-H1 and CDR-H2 are both encoded entirely by regions of the IGHV gene, whereas the CDR-H3 spans from the end of the IGHV gene to the beginning of the IGHJ gene. Similarly, the light chain CDR-L1 and CDR-L2 are both encoded in the IGKV/IGLV gene and the CDR-L3 spans the end of the IGKV/IGLV to the beginning of the IGKJ/IGLJ (see Figure 1-5).

There is a fourth heavy chain loop positioned in the FR3 region which, although not classed as a hypervariable loop, can make contact with antigen and does undergo affinity maturation. This region is therefore sometimes referred to as the fourth CDR loop (CDR-H4) (Bond et al., 2005), however this region is not as widely recognised as the six CDRs described above.

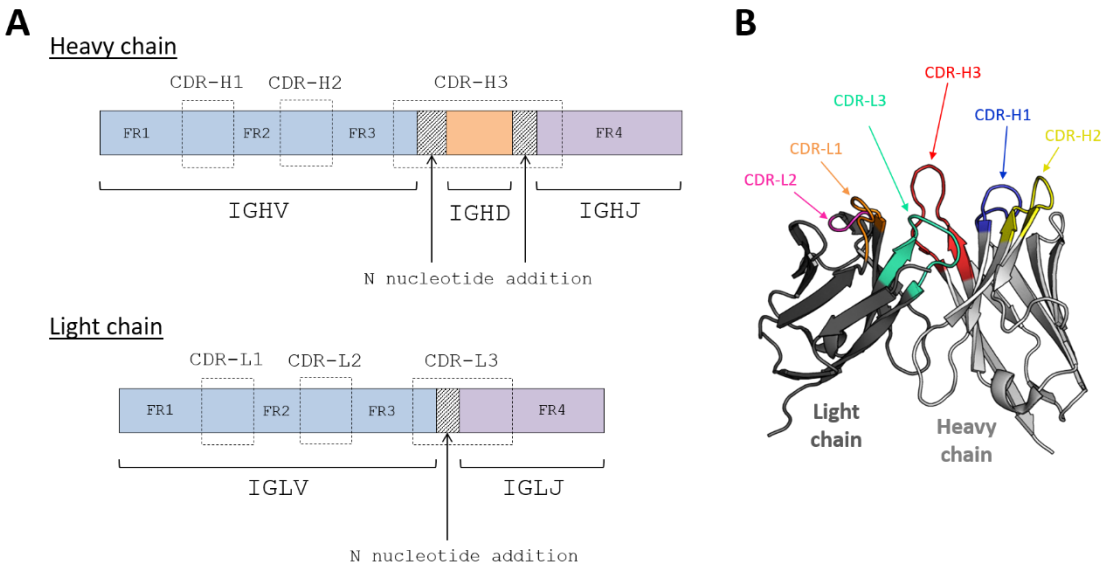


Figure 1-5: Position of CDR loops within the framework regions (FR). The CDR loops are the most diverse regions of the antibody variable region and make most contact with the antigen. The CDR-H3 is the most diverse and is thought to be the most important for antibody specificity, the CDR-L3 loop is not as diverse as there is only one region of N-nucleotide addition. The FRs are much more conserved and are primarily structural. PDB ID: 1FVC. This figure has been published in Dunn-Walters et al. (2018).

Due to the fact that the CDR-H3 loop is encoded by three separate gene segments and two regions of N-nucleotide addition (Figure 1-5A), this region is the most diverse part of the antigen-binding site and is often considered to be the most important region for antibody

specificity. Experiments in mice have shown that diversity in the CDR-H3 alone can be enough to produce an effective antibody repertoire with high levels of antibody specificity (Xu & Davis, 2000). On average, approximately 31% of energetically important residues in the antigen binding site are contributed by the CDR-H3, followed by 23% from CDR-H2. CDR-H1, CDR-L2 and CDR-L3 are approximately equally important, contributing 13% each. The contribution of CDR-L2 is usually negligible, contributing just 6% of energetically important residues (Kunik & Ofran, 2013). This means that while the majority of interactions with the antigen are formed with residues in the CDR-H3, approximately 70% of energetically important interactions are with residues on the other five loops. However, it is important to note that, as this work is based on structures in the Protein Data Bank (PDB) there is a bias towards antibodies with shorter CDR-H3.

Although the majority of antigen binding is done by the three heavy chain CDR loops (Burkovitz & Ofran, 2016), the light chain is certainly important for antigen specificity. The CDR-L3 loop is the second most diverse as it incorporates two genes and a region of N-nucleotide addition (see Figure 1-5). The light chain is important for antibody specificity and diversity in the light chain can alter the specificity of the antibody (Persson et al., 2013). This is particularly important in receptor editing where switching the light chain alters the specificity of an antibody at the central tolerance checkpoint during B cell development with the aim of rescuing autoreactive antibodies (see section 1.4, page 37).

The precise definitions of where the FRs and CDRs start and finish can vary somewhat and this can make a difference to e.g. physicochemical property calculations. Therefore, it is important to understand which numbering scheme has been used and where the FR and CDR boundaries lie.

1.3.1 Variable Region Numbering Schemes

1.3.1.1 IMGT Numbering Scheme

The IMGT numbering scheme is the most commonly used variable region numbering scheme and was first described in 1997 (Lefranc, 1997), with CDR3 and FR4 numbering added later (Lefranc et al., 2003). Importantly, the numbering scheme is the same for immunoglobulin heavy chains, light chains and also T cell receptors. In this thesis, the IMGT numbering scheme and CDR/FR definitions have been used in all cases unless stated otherwise.

In the IMGT scheme, the variable region spanning the entire V(D)J region is referred to as the V-DOMAIN and amino acid positions are numbered 1-129. Conserved amino acids are always assigned the same numbers (residues highlighted in red in Figure 1-6). FR1 is defined as the region spanning amino acids 1-26, within FR1, there is a conserved cysteine (1st-CYS) at position 23. CDR1 can be between 5 and 12 amino acids long and spans positions 27-38. FR2 contains a conserved tryptophan at position 41 (CONSERVED-TRP) and spans positions 39-55. CDR2 can be between 0 and 10 amino acids long and spans positions 56-65. FR3 spans positions 66-104, finishing with a conserved cysteine at position 104 (2nd-CYS). The CDR3 can vary a lot in length and is defined as starting at position 105 (just after 2nd-CYS) and finishing at position 117. The FR4 region begins with either a tryptophan (heavy chain) or a phenylalanine (light chain) at position 118 (J-TRP or J-PHE respectively) and ends at position 129. The region beginning with 2nd-CYS and ending with J-TRP/J-PHE (i.e. two amino acids longer than the CDR3) is known as the JUNCTION. This is followed by a conserved glycine at position 119 (Glycine 119). Glycine 119 forms a hydrogen bond with 2nd-CYS (Lefranc et al., 2003).

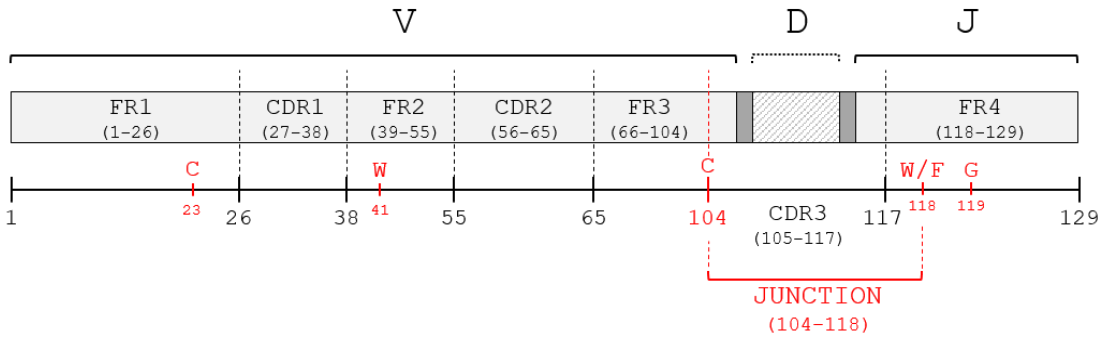


Figure 1-6: IMGT V-DOMAIN numbering scheme. This scheme applies to both antibody heavy and light chains and T cell receptors. The conserved residues are shown in red.

To allow for the large range of CDR3 lengths, the IMGT scheme employs a flexible numbering system. If a V-DOMAIN has a CDR3 of less than 13 amino acids, gaps are created by removing numbered positions symmetrically in the following order: 111, 112, 110, 113, 109, 114 etc. (Figure 1-7A). If the V-DOMAIN has a CDR3 longer than 13 amino acids, positions are added symmetrically between positions 111 and 112 in the following order: 112.1, 111.1, 112.2, 111.2 etc. (Figure 1-7B). By using this system, the CDR3 can be lengthened and shortened without affecting the numbering positions of structurally important conserved amino acids (Lefranc et al., 2003).

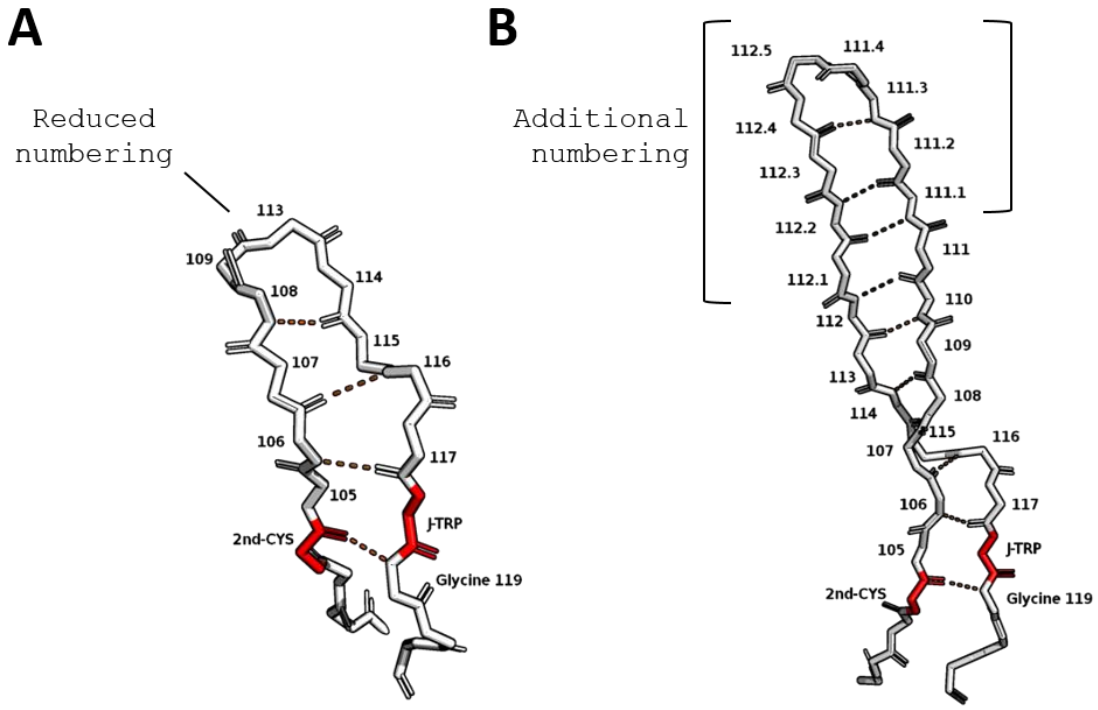


Figure 1-7: IMGT CDR3 numbering scheme. The CDR-H3 shown in (**A**) is 10 amino acids long, so numbers 111, 112 and 110 have not been used. The CDR-H3 shown in (**B**) is 22 amino acids long, so 9 amino acids have been inserted in-between positons 111 and 112. 2nd-CYS and J-TRP are highlighted in red. Polar bonds are shown in orange. The CDR-H3 loops in A and B are from PDB structures 4KQ3 and 3GHE respectively.

1.3.1.2 Kabat and Chothia Numbering Schemes

The Kabat numbering scheme was the first standardised numbering scheme to be created (Kabat et al., 1983) and the improved Chothia numbering scheme followed soon after (Chothia & Lesk, 1987). The Kabat numbering scheme is more sequence-focussed, whereas the Chothia scheme is more structure-focussed (Abhinandan & Martin, 2008) and takes spatial positioning into account (corrected CDR-L1 and CDR-H1 (Honegger & Plückthun, 2001)). The Chothia CDR-H1 and CDR-L1 numbering schemes changed slightly in 1989 (Chothia et al., 1989), but reverted in 1997 (Al-Lazikani et al., 1997).

The Chothia numbering scheme is popular among structural modellers due to the straightforward way in which it deals with indels (Kovaltsuk et al., 2017). In this thesis, the Chothia scheme has only been used when modelling CDRs.

Unlike the IMGT numbering scheme, the Chothia scheme has separate numbering for heavy chains and light chains (Figure 1-8). This scheme was devised using only the structure of the CDR loops, therefore any structurally important residues in the framework regions was not considered (Abhinandan & Martin, 2008). The CDR-H1 region occurs between residues 25 (three residues upstream of the conserved cysteine) and 32, additional amino acids are inserted in-between residues 31 and 32 using 31A and 31B. The CDR-H2 region occurs between residues 51 and 56. Longer loops are accommodated by additional numbering between residues 52 and 53 using 52A, 52B and 52C. The CDR-H3 is located between residues 94 (two residues upstream of the conserved cysteine) and finishes at position 102 (one residue downstream of the conserved tryptophan). Additional amino acids are added between residues 100 and 101 using lettering 100A-100K. The CDR-L1 region occurs between residue 23 (conserved cysteine) and 32. Additional amino acids are added between residues 30 and 31 with the lettering 30A-30F. The CDR-L2 region occurs between residues 49 and 56, there is no provision for additional amino acids The CDR-L3 region occurs between residue 88 (conserved cysteine) and 97. Additional amino acids are added between residues 95 and 96 using the lettering 95A-95F (Abhinandan & Martin, 2008).

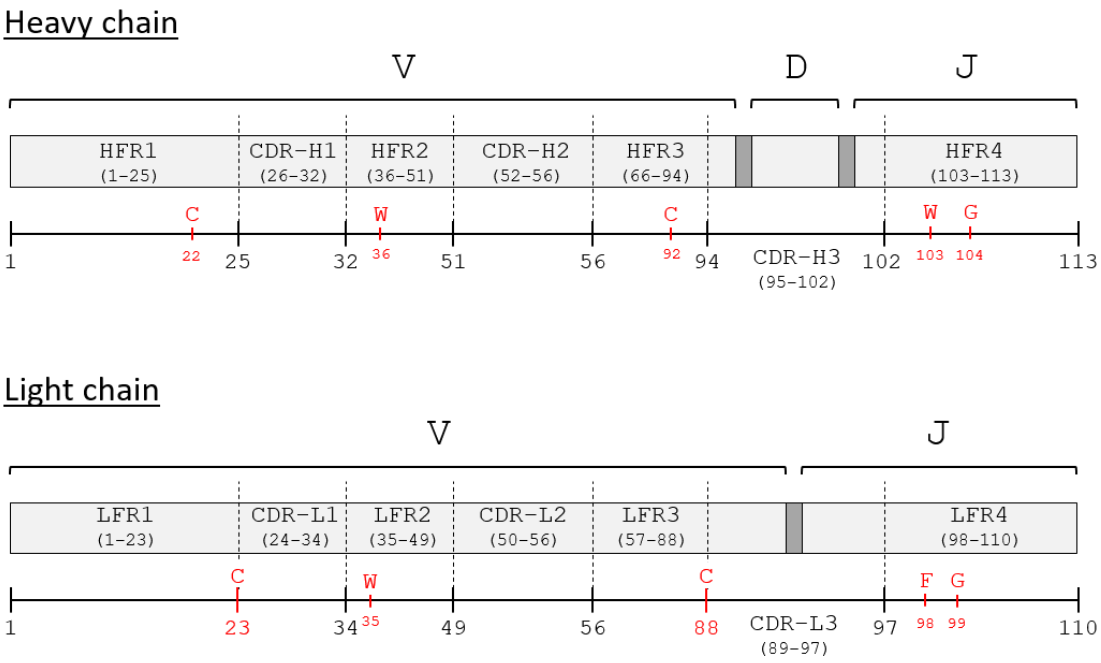


Figure 1-8: Chothia numbering scheme. There is a separate scheme for antibody heavy and light chain variable domains and it does not apply to T cell receptors.

1.3.1.3 Other Numbering Schemes

A number of other numbering schemes have been developed. The AHo numbering scheme was described in 2001 and was the first scheme to use a numbering method which allowed CDR loops to extend symmetrically around key positions (Honegger & Plückthun, 2001). The Martin numbering scheme, an enhanced Chothia numbering scheme, was described in 2008. This was developed to address errors in the Chothia numbering scheme, particularly in the numbering of CDR-H2 and heavy chain FR3, and addressed the fact that the original Chothia scheme did not consider indels in the framework regions (Abhinandan & Martin, 2008). A paper published by North et al. (2011) did not propose a new numbering scheme, but after a comprehensive review of CDR canonical structures, did use CDR definitions which differed slightly from Kabat and Chothia.

1.4 B cell Development and Central Tolerance

V(D)J recombination and receptor editing happens as B cells develop in the bone marrow. Due to the vast potential diversity created by V(D)J recombination, it is possible to get variable regions which bind self-antigen (autoreactive antibodies). If released from the bone marrow into the peripheral blood, these antibodies could cause autoimmune disease. Developing B cells in the bone marrow do not secrete antibody, instead they express their antibodies on the cell surface as B cell receptors (BCRs). At each stage of development, B cells are subject to positive and negative selection pressures to ensure that the B cell is developing correctly and is expressing functional BCRs which are not autoreactive. B cells which do express autoreactive BCRs are selected out of the repertoire during a process called central tolerance.

Figure 1-9 shows the stages of B cell development in the bone marrow. VDJ recombination of the IgH locus happens in pro-B cells and the recombined heavy chain is then displayed on the surface of large pre-B cells in combination with a surrogate light chain. This structure is known as a pre-BCR and its purpose is to check that the heavy chain is functional and capable of pairing with light chains. If the heavy chain does not pair successfully with the surrogate light chain, a further heavy chain rearrangement is made using the second heavy chain allele. If the second heavy chain rearrangement also cannot pair, the B cell dies. However, if the heavy chain does pair successfully with the surrogate light chain, the B cell progresses to the small pre-B cell stage and the light chain loci undergo VJ recombination. The paired heavy and light chains are then displayed on the immature B cell surface as a BCR and it is at this point that central tolerance acts (Lydyard et al., 2011).

During central tolerance, BCRs are tested for binding against antigens in the bone marrow. BCRs that do not bind any antigen in the bone marrow progress through the immature B cell stage and are released from the bone marrow into the peripheral blood. However, it is thought that if a BCR binds antigen in the bone marrow, the BCR is autoreactive and therefore should be removed from the repertoire. If a functional BCR on an immature B cell binds self-antigen in the bone marrow, there are four possible outcomes for the B cell (Bashford-Rogers et al., 2018):

1. Deletion
2. Anergy
3. Receptor editing
4. Ignorance

Clonal deletion and anergy both result in the removal of the functional B cell from the repertoire. Deletion uses B cell death to remove autoreactive B cells (Nemazee & Bürki, 1989). Anergy, on the other hand, is when the B cell enters a state of unresponsiveness in which the B cell is not stimulated by antigen binding (Goodnow et al., 1988). However, the primary mechanism of central tolerance is a process called receptor editing. In receptor editing, the specificity of the BCR is altered by further rearrangement of the light chain loci (and occasionally the heavy chain loci, although this is rare) (Nemazee, 2006). Kappa light chain rearrangement tends to precede lambda and lambda light chains seem to be better at rescuing autoreactive heavy chains than kappa (Wardemann et al., 2004). Light chains with low CDR isoelectric points (pI) or high aspartic acid usage are good at rescuing DNA-reactive autoantibodies (Li et al., 2001, Kalinina et al., 2014). If the autoreactive BCR does not encounter its cognate antigen in the bone marrow, the B cell will mature as if the BCR were not autoreactive; this is known as immunological ignorance.

It is not well understood exactly under which circumstances a B cell will undergo receptor editing, or enter anergy or deletion. However, it is thought that B cells which bind in a strongly cross-linking manner to multivalent self-antigen are deleted from the repertoire (Fulcher et al., 1996), whereas anergy is thought to be used when the B cell encounters weakly cross-linking antigen of low valence (Hartley et al., 1991). If receptor editing does not rescue an autoreactive BCR, the B cell is removed from the repertoire by deletion (Tiegs et al., 1993).

It is thought that 55-75% of early immature B cells are autoreactive and the majority of these are removed from the repertoire by central tolerance (Wardemann et al., 2003).

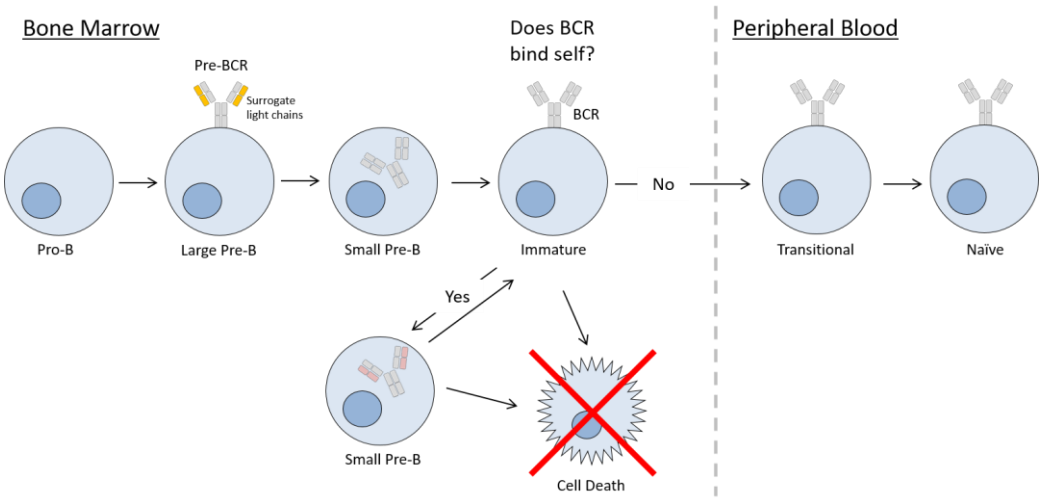


Figure 1-9: Central tolerance and receptor editing. Central tolerance acts on immature B cells to determine if the BCR binds self-antigen and is therefore autoreactive. If the BCR does not bind self-antigen, the immature B cell matures and is released into the peripheral blood. If the BCR does bind self-antigen, the BCR is pulled back inside the B cell and undergoes receptor editing. If receptor editing cannot rescue the B cell, it is removed from the repertoire by apoptosis. Diagram modified from Nemazee (2006) and Nemazee (2017).

1.5 B cell Activation and Peripheral Tolerance

1.5.1 T-Independent B cell Activation

T-independent B cell activation is triggered by certain types of antigen, known as thymus-independent antigen (TI antigen). There are two types of TI antigen: TI-1 antigen and TI-2 antigen.

TI-1 antigen engages toll-like receptors (TLRs) on B cells in addition to BCRs in order to trigger B cell activation. Common TI-1 antigens include lipopolysaccharide (LPS), bacterial DNA and some viral particles (Fehr et al., 1998). B cells activated by TI-1 antigen only produce IgM antibodies and do not undergo somatic hypermutation or class switching. Although this limits the potency of the response, it is still a useful first line of defence.

TI-2 antigens are large, highly repetitive carbohydrate or protein epitopes on the surface of pathogens. The large amount of TI-2 antigen on the pathogen surface cross-links multiple BCRs forming a cluster of cross-linked BCRs. This high level of cross linking triggers a Bruton's tyrosine kinase (Btk)-dependent signalling cascade which overrides the requirement for a secondary activation signal and the B cell is activated in the absence of T cell help or TLRs (Vos et al., 2000). It was thought that TI-2 antigens only elicited primary antibody responses which did not lead to the generation of memory B cells. However, it has been shown that TI-2 antigens can elicit a memory response against polysaccharides if aided by antigen-specific IgG antibodies (Obukhanych & Nussenzweig, 2006). Somatic hypermutation and class switching can also occur at low levels in response to TI-2 antigens (Toellner et al., 2002).

1.5.2 T-Dependent B cell Activation

Other antigens (primarily protein antigens) stimulate a T-dependent response, which means help from CD4+ T cells is required to activate the B cell against that antigen. When a B cell recognises a cognate T-dependent antigen in the lymphoid follicles, the B cell internalises the antigen and displays the processed peptide on its MHC class II molecules. In the T cell zone of the lymphoid follicle, CD4+ T cells bearing complementary TCRs bind the antigen-MHC class II complex on the B cell. The CD40L on the T cell also binds the CD40 receptor on the B cell and the CD28 co-receptor on the T cell binds CD80 and CD86 on the B cell (Owen et al., 2013, Hoffman et al., 2016). The T cell-activated B cells then either form primary foci, or germinal centres.

The B cells which form primary foci do so at the interface between the T cell zone and B cell zone of the lymphoid follicle. These T cell-activated B cells differentiate into plasma cells and then migrate to the medullary cord regions of the lymph nodes or the red pulp region of the

spleen where they secrete large amounts of antigen-specific IgM antibodies. Many of these cells die off after the primary response, but some migrate to the bone marrow where they remain as long-lived plasma cells. Antibodies produced by B cells in the primary foci provide a useful early antibody response as they appear in circulation approximately four days after their initial antigen stimulation (Owen et al., 2013).

In contrast, the B cells which form the germinal centres in the lymph nodes and spleen take longer to mature (they appear in circulation approximately 6-10 days after their initial antigen stimulation) however, they produce a much more specific antibody response. Further differentiation of the germinal centre B cells is aided by follicular helper T cells (T_{FH}) and follicular dendritic cells (FDC). Within the germinal centres, B cells undergo intense proliferation and affinity maturation. As B cells proliferate, further diversity is introduced into the BCR variable regions via somatic hypermutation (SHM) at a rate of approximately one mutation per 1000 bp per division. The B cells where SHM has resulted in increased affinity for cognate antigen will out-compete B cells bearing lower affinity BCRs and be preferentially selected. Alongside SHM, class switch recombination (CSR) occurs. CSR involves the recombination of the constant region genes, thus changing the functionality of the antibody. B cells stimulated in a T-dependent manner become plasma cells and memory cells (Owen et al., 2013). Both SHM and CSR are directed by activation-induced cytidine deaminase (AID) and are described in more detail in sections 1.6 and 1.7 respectively.

1.5.3 Peripheral B cell Tolerance

Autoreactive B cells are removed from the repertoire during central tolerance using apoptosis, anergy or receptor editing. However, central tolerance does not remove all autoreactive B cells and some escape into the periphery, perhaps due to immunological ignorance. Therefore a second tolerance checkpoint, which is believed to act primarily on the transitional B cells (Figure 1-9, page 38), removes autoreactive B cells in the peripheral blood (Wardemann et al., 2003). This checkpoint is known as peripheral tolerance.

Peripheral tolerance is thought to work as the inverse of T-independent and T-dependent B cell activation. If a BCR binds an antigen but the B cell does not receive the necessary secondary signals, the B cell is removed from the repertoire by apoptosis or anergy. Peripheral tolerance is necessary because autoreactive B cells may escape central tolerance or they may be created during somatic hypermutation. As autoreactive T cells are removed from the T cell repertoire during thymic maturation, the chances of a T cell being present which is also autoreactive to the same autoantigen is low (Parham, 2009).

1.6 Somatic Hypermutation

Once a B cell has been successfully activated by its complementary antigen, the binding efficacy of the antibody is increased by somatic hypermutation (SHM). During SHM, point mutations are introduced to the heavy and light chain variable regions by activation-induced cytidine deaminase (AID).

AID is expressed only in proliferating B cells and it converts cytosine to uracil in ssDNA, which is temporarily available during transcription. Uracil is a normal component of RNA, but not DNA. However, dU mimics dT during replication and therefore, although the cellular DNA repair pathways can convert the uracil into any of the four DNA bases, the dU tends to be converted into dT, resulting in a bias towards C→T and G→A mutations (Maul & Gearhart, 2010). The rate of mutation in SHM is one variable region mutation per cell division (this is over a million times greater than the mutation rate of a normal gene) (Parham, 2009).

SHM occurs in the germinal centres and some mutations will result in increased affinity for the target antigen, others will result in decreased affinity. This leads to a Darwinian-like selection in which somatically mutated B cells must compete for a limited supply of antigen presented by follicular dendritic cells and helper T cells. The B cells bearing BCRs with the highest affinity for the antigen compete more effectively for antigen and are therefore preferentially selected to mature into plasma cells and memory cells (Victora & Nussenzweig, 2012). This process is known as affinity maturation. Mutations are concentrated in the CDRs rather than the FRs as CDRs have more contact with antigen (Parham, 2009).

In addition to altering the protein composition of an antibody variable region, N-glycosylation sites can be introduced and removed in the antibody variable region during SHM and this also affects antibody specificity (Dunn-Walters et al., 2000). Despite there being only five germline genes which encode an N-linked glycosylation site (IGHV1-8, IGHV4-34, IGHV5-10-1, IGLV3-12 and IGLV5-37), antibodies with N-glycans in their Fab regions are present at a frequency of about 15-25% of serum IgG antibodies, suggesting that these sites are commonly introduced during SHM (van de Bovenkamp et al., 2016).

1.7 Class Switch Recombination

1.7.1 Mechanism of Class Switch Recombination

Class switch recombination (CSR) is the process by which B cells switch from expressing antibodies using IgM/IgD constant regions to antibodies using IgA, IgE or IgG constant regions. CSR occurs almost exclusively as a result of T-dependent activation and it occurs alongside affinity maturation in the germinal centres (however some CSR can occur in a T-independent manner (Berkowska et al., 2011)). By swapping the constant region used by the antibody, the function of the antibody changes. Different antigens will elicit different constant region usage. CSR is difficult to study *in vivo* and so the precise triggers inducing the process is not fully understood, however, it is thought that it is triggered by BCR stimulation in combination with signals from CD40 and TLRs (Stavnezer & Schrader, 2014).

Like SHM, CSR is mediated by AID. AID mediates the recombination between switch regions (guanine-rich regions with a high frequency of A/T-G-C-A/T motifs which sit just upstream of each constant region gene) by converting cytosines to uracils in an upstream donor switch region and a downstream acceptor switch region. Enzymes in the base excision repair pathway and the mismatch repair pathway then convert the uracils into double-stranded DNA breaks. The two DNA ends are then brought together by non-homologous end joining (Stavnezer & Schrader, 2014). The DNA between the donor and acceptor switch regions is released as a loop of DNA (Figure 1-10A). Successive CSR events can occur (Gearhart et al., 1980, Zhang et al., 1994, Jackson et al., 2014), although only in the downstream direction as the intermediate DNA is looped out with each recombination. Antibodies using more downstream constant regions have higher levels of SHM. This is thought to be due to prolonged exposure to AID in the germinal centres (Jackson et al., 2014). Every CH gene can undergo alternative splicing to generate two carboxy termini: one anchors the immunoglobulin on the B cell as a BCR, the other allows the immunoglobulin to be secreted as an antibody (Schroeder & Cavacini, 2010).

Certain recombination events are more common than others, for example, IgM/IgD B cells are most likely to switch to either IgG1 or IgA1 than other subclasses (Figure 1-10B). Subclasses are more likely to switch between members of the same class than switch to a different class (Kitaura et al., 2017). The data presented in Kitaura et al. (2017) suggests that the majority of expanded clones are of a single class, approximately 9% of expanded clones contain a mixture of subclasses within the same class, and only a very small proportion (approximately 1%) of expanded clones contain more than one class. However, this study used a very stringent definition of a clone (identical CDR-H3) and so the proportions may differ if a different definition was used.

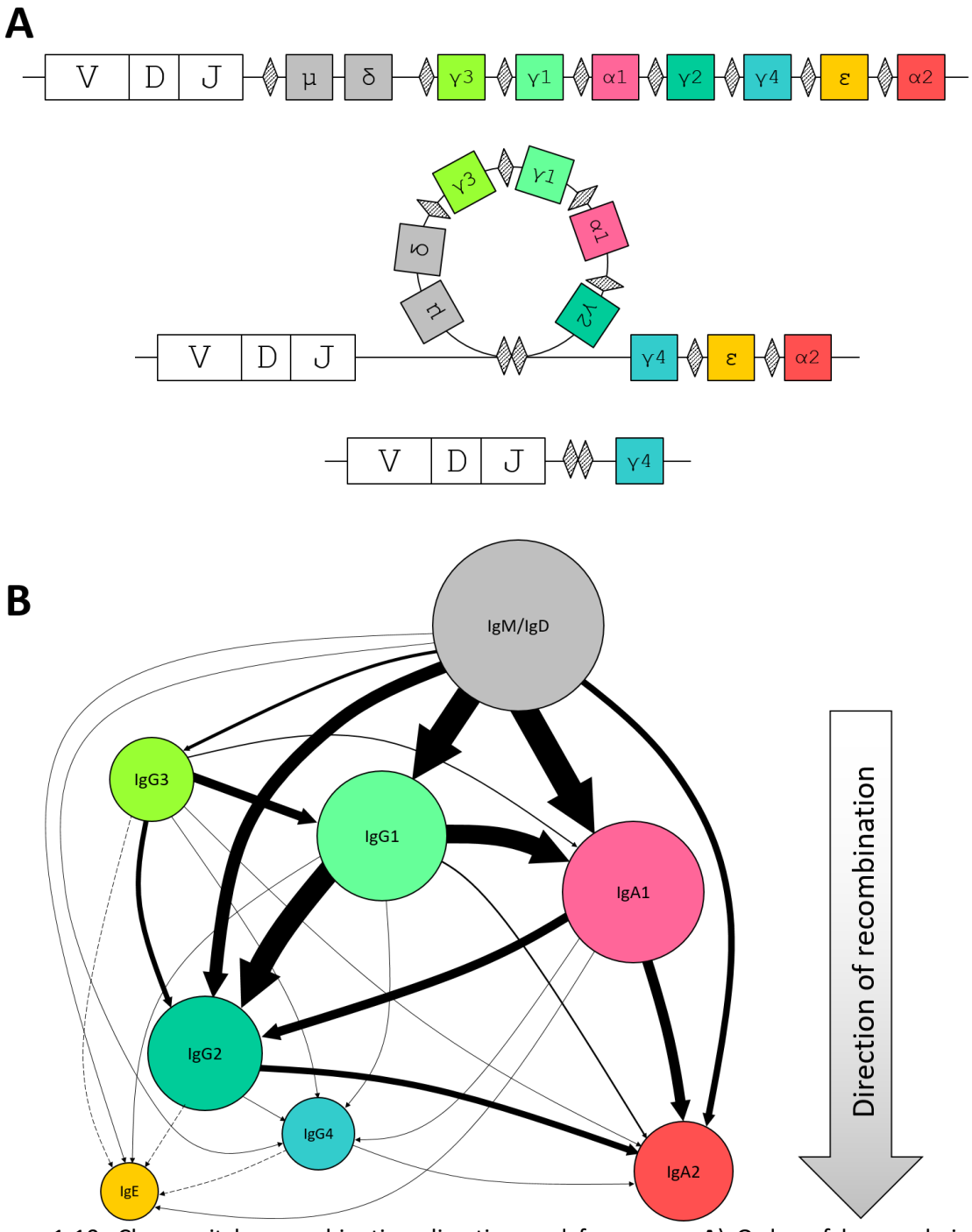


Figure 1-10: Class switch recombination direction and frequency. **A**) Order of heavy chain constant region (CH) genes in the genome. When class switch recombination occurs, the intermediate section of DNA gets looped out and discarded. Therefore further recombination can only occur in the downstream direction. Diamonds represent switch regions. Diagram adapted from Stavnezer et al. (2008). **B**) Recombination occurs at different frequencies. The most common switch for IgM/IgD expressing B cells is to IgG1 or IgA1. IgG1 then more commonly switches to other IgG subtypes and IgA1 most often switches to IgA2. Circle diameter reflects subtype abundance in the repertoire and arrow width is proportional to the frequency of each class switch as reported in Horns et al. (2016). Dotted lines represent class switches which are so uncommon they were not observed in the Horns et al. (2016) dataset. Diagram adapted from Horns et al. (2016).

1.7.2 Effector Functions of Heavy Chain Classes

The five heavy chain classes of antibody are IgM, IgD, IgG, IgA and IgE. IgG and IgA are further subdivided into four and two subclasses respectively. The constant regions of IgG, IgA and IgD consist of three domains (CH1-CH2-CH3), whereas IgM and IgE have an additional domain (CH1-CH2-CH3-CH4). CH1 is located in the Fab region, whereas the remaining CH comprise the Fc region (Figure 1-11). The Fc fragments of the different heavy chain classes have different effector functions and CSR facilitates the selection of the most appropriate Fc for each immune response.

The Fc regions mediate antibody effector functions by binding to Fc receptors on immune cells to activate responses such as ADCC and ADCP or by initiating complement. Different classes dominate in different reactions. For example, IgG is the most common class in the blood, whereas secreted IgA is most prevalent at mucosal surfaces. IgE is generally used in response to extracellular parasites and in allergic reactions (Schroeder & Cavacini, 2010).

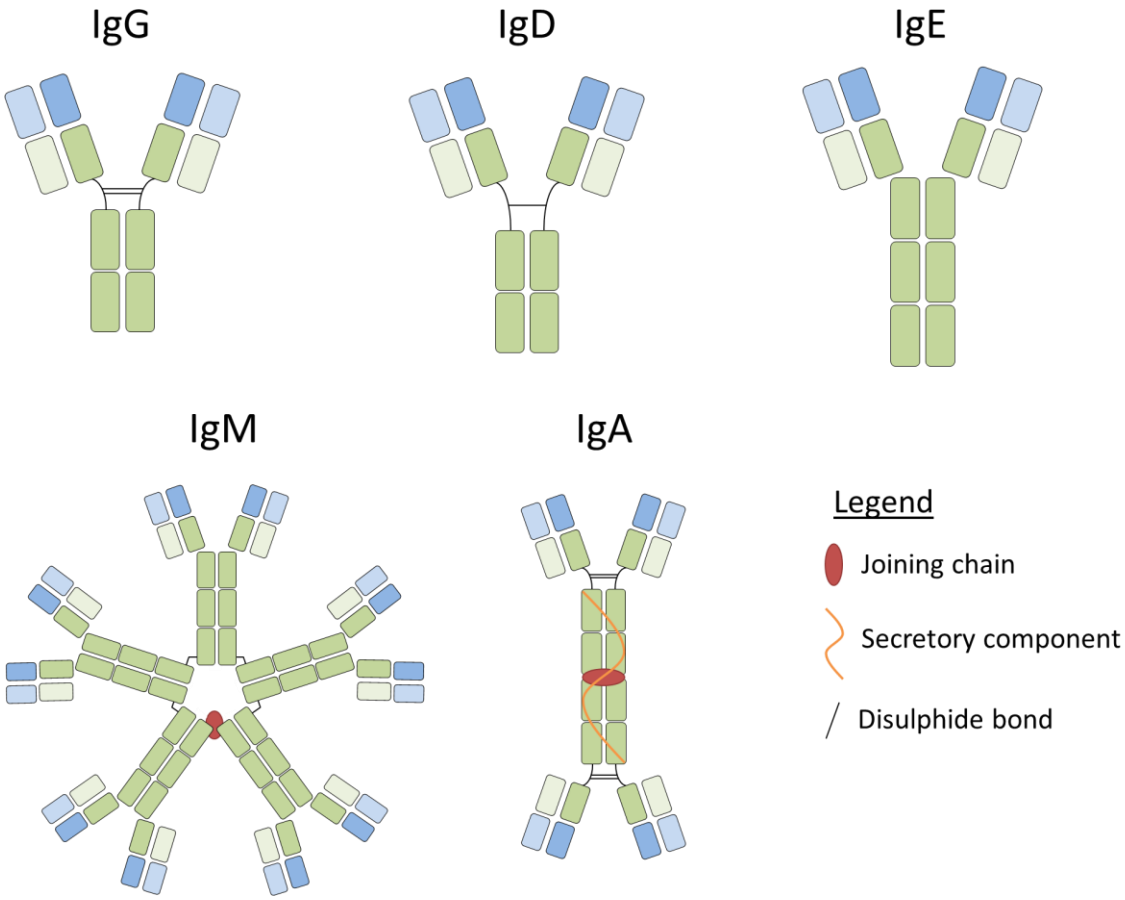


Figure 1-11: Antibody heavy chain classes form different conformations. The different Fc regions have differing numbers of CH domains, different glycosylation patterns and form different structures (monomers, dimers and pentamers). Each class is best suited to different roles in the immune response. The secretory component of IgA is only present on the secreted form of IgA.

1.7.2.1 IgM

Monomeric IgM is displayed on the surface of all developing and antigen-naïve B cells in the form of B cell receptors (BCRs). Upon binding their specific antigen, BCR signalling triggers the maturation of the B cell. In addition to the membrane-bound IgM, there are two types of secreted IgM: natural IgM and adaptive IgM.

Natural IgM arise spontaneously without antigenic stimulation. Natural antibodies provide an important innate-like first line of defence against invading pathogens and comprise a substantial part of peripheral blood lymphocytes in newborns. Natural antibodies often have near-germline V(D)J configurations and they bind antigen with a lower affinity than affinity-matured antibodies. They are commonly polyreactive (Zhou et al., 2007). In addition to binding pathogens and activating complement, natural antibodies can also clear dead or dying host cells and exhibit anti-inflammatory effects (Grönwall & Silverman, 2014).

Adaptive IgM is affinity matured and therefore has a high affinity. It was thought that IgM antibodies were not somatically mutated and could not form a memory response. However, it has been shown that they can be somatically mutated and IgM memory cells do exist (Díaz-Zaragoza et al., 2015).

IgM antibodies often have low affinities as they are often not somatically mutated. However, secreted IgM forms pentamer structures with ten antigen-binding sites resulting in high avidity. IgM pentamers are held together by disulphide bonds in the CH4 regions and incorporate a joining chain (J-chain) which facilitates secretion at mucosal surfaces (Schroeder & Cavacini, 2010). IgM pentamers are a very efficient activator of complement; a single IgM pentamer bound to an antigen is sufficient to bind the first complement component (C1) and start the complement cascade (two IgG in close proximity are alternatively required). Complement results in activation of mast cells, attraction of neutrophils, enhanced opsonisation by phagocytes and, if the pathogen is a microbe, lysis via the membrane attack complex (Lydyard et al., 2011).

1.7.2.2 IgD

IgD is co-expressed as a monomer alongside IgM on the surface of naïve B cells. However its role in the immune response is largely unknown. IgD is downregulated once the B cell has encountered antigen, however, B cells do occasionally class switch to IgD. These class-switched IgD antibodies are usually paired with lambda light chains (Arpin et al., 1998). IgD-secreting plasma cells are primarily found in the upper respiratory tract such as the nasal mucosa and

the tonsils. B cells appear to undergo IgD CSR *in situ* and the resulting IgD antibodies recognise respiratory bacteria and bind circulating basophils (Chen et al., 2009).

1.7.2.3 IgG

IgG is the most abundant antibody class in human serum. This class is composed of four subclasses: IgG1, IgG2, IgG3 and IgG4 (named in order of decreasing abundance). These four subclasses of constant region genes share over 90% of their amino acid sequence identity, however there are significant structural differences in the hinge region (Liu & May, 2012) and upper CH2 domain, both of which are important for IgG binding to Fc-gamma receptors (FcγR) and complement component 1q (C1q) which elicit responses such as antibody-dependent cellular cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP), and complement-dependent cytotoxicity (CDC) (Irani et al., 2015, Chan et al., 2015). The differences in Fc regions mean that each subclass predominantly targets certain types of antigen and is suited to certain effector functions. Each subclass also has a number of allotypes (constant region allelic variants) which affect the effector functions.

IgG1 and IgG3 have very similar roles; both subtypes primarily react to soluble and membrane-bound proteins. Anti-viral responses are mainly IgG1/IgG3, for example HIV (Raux et al., 2000) and Ebola (Leroy et al., 2001). The antibody response against *Plasmodium falciparum* malaria is also IgG1/IgG3 (Weaver et al., 2016). IgG1 and IgG3 Fc bind C1q, initiating complement and they also both bind all of the FcγR, indicating that they initiate cell-mediated responses such as ADCC (Vidarsson et al., 2014). Although IgG1 and IgG3 have similar effector functions, an IgG3-dominant response is rare. Further, almost all therapeutic antibodies are of the IgG1 isotype and there are no approved IgG3 therapeutic antibodies. This may be due to the fact that IgG1 is the most prevalent subtype in human serum, or it may be because IgG3 has a very extensive hinge region which makes it vulnerable to proteolysis. More practically, it may be because IgG3 has a short half-life and it cannot be purified using protein A (Irani et al., 2015).

IgG2 binds primarily to polysaccharides and is almost exclusively responsible for the IgG responses to bacterial capsular polysaccharide. People deficient for IgG2 antibodies experience recurrent bacterial infections, for example *Streptococcus pneumoniae* (Barrett & Ayoub, 1986). IgG2 is not a potent inducer of complement and only binds certain FcγRIIa allelic variants (e.g. H131). Therefore people with the H131 FcγRIIa allotype have better IgG2 opsonisation of bacteria such as *Streptococcus pneumoniae* and greater neutrophil phagocytosis (Jefferis, 2012).

IgG4 primarily binds protein antigen and is produced as the result of a Th2-dependent response. Like IgE, IgG4 is often raised in response to allergens and has been found to be important in the antibody response against *Schistosoma mansoni* (Boctor & Peter, 1990). IgG4 does not induce complement and is not a potent binder of most FcγR. IgG4 responses are often formed following long-term or repeated exposure to an antigen, for example after a course of immunotherapy. It is thought that a major role of IgG4 is to regulate inflammatory IgE responses by blocking IgE antibodies from binding the antigen (Nouri-Aria et al., 2004, Francis et al., 2008, James et al., 2011, Boctor & Peter, 1990).

IgG4 antibodies display a phenomenon known as IgG4 Fab arm exchange in which IgG4 half-molecules randomly recombine with each other creating a bispecific antibody which is monovalent for each antigen; these antibodies cannot form immune complexes (Aalberse et al., 2009, Nirula et al., 2011). It is thought that the rate at which Fab arm exchange takes place is controlled by the redox environment and therefore may vary between tissues (Rispens et al., 2011).

1.7.2.4 IgA

Secretory IgA is the predominant antibody at most mucosal surfaces, and serum IgA is the second most prevalent blood antibody after IgG. Serum IgA is made in the lymph nodes and bone marrow and tends to be monomeric, whereas secretory IgA is made in the lymphoid tissues at mucosal surfaces and tends to be dimeric. There are two IgA subclasses (IgA1 and IgA2), approximately 90% of serum IgA is IgA1, whereas secretory IgA is approximately 60% IgA2 and 40% IgA1, however this ratio varies between mucosal locations (Woof, 2013).

The major structural difference between IgA1 and IgA2 is that IgA1 has a much longer hinge region than IgA2. This may give IgA1 a higher avidity by allowing bivalent reactions with more distant antigens, however the extended hinge region does make IgA1 vulnerable to cleavage by proteases expressed by certain pathogenic bacteria (Woof & Kerr, 2004).

IgA can activate both the alternative and the lectin complement pathways (Roos et al., 2001).

IgA interacts with the Fc-alpha receptor FcαRI which is expressed on cells such as neutrophils, eosinophils, monocytes, macrophages and Kupffer cells. Both monomeric and dimeric IgA is able to interact with FcαRI, however secreted IgA binding is hampered by steric hindrance caused by the secretory component. Therefore, while serum IgA is capable of triggering inflammatory responses such as phagocytosis and ADCC, secretory IgA is thought to be non-inflammatory and primarily plays a role in preventing the adherence of microorganisms to mucosal surfaces (Bakema & van Egmond, 2011). Additionally, secretory IgA is present at very

high concentrations in breast milk and plays an important role in the passive protection against infection of infants (Hanson & Korotkova, 2002).

Unlike the blood antibody repertoire, which have a bias towards kappa light chains, secreted antibodies have a bias towards lambda light chains (Mole et al., 1994).

1.7.2.5 IgE

IgE is the least prevalent class in the blood, with a half-life of just three days (compared to the IgG half-life of 20 days). However, most IgE is not in the serum, but is sequestered in relevant tissues (Gould et al., 2003).

IgE is produced as part of a Th2 response against parasitic worm infections such as *Schistosoma mansoni* and *Trichinella spiralis*. IgE Fc interacts with two Fc-epsilon receptors: FcεRI and CD23 (FcεRII). FcεRI is a high-affinity IgE receptor and is expressed on mast cells, basophils, antigen-presenting cells, monocytes and eosinophils. The binding of FcεRI leads to degranulation of mast cells and a strong inflammatory response. IgE plays a role in allergy when these responses are raised against something harmless such as pollen (Gould & Sutton, 2008).

Certain IGHV families are associated with certain allergies. For example, there is a bias towards IGHV1 usage in the IgE antibodies of people with peanut allergies (Janezic et al., 1998). IGHV4-21 is prevalent in autoantibodies against red blood cells as a result of Epstein-Barr Virus infection (Chapman et al., 1993). IGHV5 has been associated with atopic dermatitis (Van der Stoep et al., 1993), asthma (Snow et al., 1995) and allergic rhinitis (Coker et al., 2005).

1.8 High Throughput Sequencing of Immunoglobulin Repertoires

High-throughput sequencing technologies have become more accurate and more affordable in recent years. With this advance in technology, high-throughput sequencing has become a more popular method of investigating the B cell repertoires of healthy individuals and various patient groups. A summary of the main stages involved in building high-throughput B cell repertoire sequencing libraries is given below.

1.8.1 Isolating Functional Antibody Variable Region Sequences

The initial step of building a B cell repertoire library for high-throughput sequencing (HTS) is the isolation of functional immunoglobulin variable region sequences. These can be isolated from B cell gDNA or mRNA obtained from sorted B cell populations or total RNA from whole blood/tissue. If mRNA is used, this is reverse transcribed either using oligo-dT primers, random primers or specific primers. The immunoglobulin transcripts are amplified using PCR.

There are two drawbacks of using PCR to generate enough sample material for HTS. Firstly, even if a very high-fidelity DNA polymerase is used, PCR does not have an error rate of zero. As immunoglobulin variable regions are extremely diverse (due to combinatorial diversity, junctional diversity and somatic hypermutation), it is often impossible to tell if a sequence contains errors. However, the number of errors can be minimised by using a high fidelity DNA polymerase and few rounds of PCR. The second drawback of using PCR to amplify variable regions is bias introduced by primers. It is common to use a set of V gene-specific and J gene-specific primers based on the set described in Sblattero &Bradbury (1998). However, some primers may be present in the mix in higher quantities than others and some primers may bind better than others leading to biases in the frequency with which certain sequences are amplified and therefore represented in the resulting dataset. Furthermore, if there are mutations at the beginning of the V gene or the end of the J gene (due to either allelic variation or somatic hypermutation) these are erased by the primer sequence and this information is lost from the resulting dataset. A better method of variable region amplification is to introduce universal primer site for example using the 5'-template switch method described in Mamedov et al. (2013). In this method, a universal primer site is added at the 5' end of the cDNA transcript during reverse transcription. A single forward primer which binds the universal primer site is then used for all amplifications.

The template switch method has the added advantage that it allows the introduction of unique identifiers (UIDs) to the cDNA transcripts. Later, in the bioinformatic analysis, UIDs enable distinction between PCR amplification (identical UIDs) and clonal expansion (similar/identical variable regions with different UIDs). The mixture of UIDs needs to be diverse enough that the

chances of two variable regions being tagged with the same UID is low – this is known as the birthday paradox (Friedensohn et al., 2017). However, although UIDs allow the distinction between mRNA transcripts, this does not necessarily mean that the mRNAs came from different cells. Different cell types express different amounts of antibody, for example, plasma cells express large amounts of antibody and therefore have large amounts of immunoglobulin mRNA. One way to distinguish between mRNA originating from different cells is to sort the cells with barcoded beads using microfluidics. This method also addresses the problem of maintaining native heavy-light chain pairing information.

Maintaining native heavy-light chain pairing is a significant problem in B cell repertoire analysis. IgH, IgK and IgL loci are located on separate chromosomes and are transcribed on separate mRNA. A B cell expressing functional immunoglobulins will express heavy chain mRNA and light chain mRNA on separate transcripts. Therefore, if all the mRNA from whole blood or a population of sorted B cells is used to amplify variable region sequences in a single reaction, it is impossible to determine which heavy and light chain sequences were paired in the individual B cells. One method sometimes employed is to assume that paired heavy and light chain sequences occur at similar frequencies in the repertoire and so heavy-light chain pairs can be inferred by matching frequencies. However, this is not a reliable method as primer biases can affect the extent to which sequences are amplified and repertoires tend to follow power-law distributions (Friedensohn et al., 2017). A “low-throughput” method of maintaining heavy-light chain pairing information is to sort single B cells into individual wells of 96-well plates and perform PCR reactions in the single wells (as used in Wardemann et al. (2003), for example). This method works as a way of maintaining heavy-light chain pairing information, however it is not possible to create very large datasets due to the large amount of time and reagents it would involve. So far, the most promising high-throughput method of maintaining heavy-light chain pairing information is encapsulating individual B cells in oil emulsion droplets containing oligo-dT-coated magnetic beads. The cell is lysed within the droplet and the mRNA is captured on the bead. The beads are then re-encapsulated in droplets containing reverse transcription and PCR reagents and overlap-extension PCR can take place, joining the heavy and light chain transcripts with a linker sequence. If the oligo-dT beads are also labelled with UIDs, and a single bead is encapsulated with a single cell, the transcripts are also labelled with a cell-identifying UID (as opposed to mRNA-identifying UIDs) and clonal expansions can be identified with even more certainty. A pioneer of this method was George Georgiou (DeKosky et al., 2015, McDaniel et al., 2016) and the technique is beginning to become more widespread (Rajan et al., 2018). However, encapsulating single cells in droplets and performing PCR within these droplets is technically challenging.

A further consideration is that the B cell repertoire (repertoire of functional rearranged immunoglobulin sequences encoded on mRNA in circulating B cells) may not be the same as the antibody repertoire (the physical antibody proteins present in the serum). Antibodies may be expressed from memory B cells in the bone marrow, and many B cells circulating in the blood will not be expressing soluble antibodies. It seems that only a small fraction of the antibodies encoded in circulating B cells express antibodies in the serum (Lavinder et al., 2015). With the use of mass spectrometry, the antibody repertoire is something which can now be investigated.

1.8.2 Clonality Analysis

There are many definitions of clonal expansions with the very simplest being identical VDJ gene usage and identical CDR-H3 sequences. It is generally agreed that a clonal expansion should use the same V genes and J genes. Some papers also require identical CDR-H3 amino acid sequences. Although this may be suitable for the analysis of antigen-naïve repertoires, the stipulation for identical CDR-H3 is unnecessarily harsh when analysing antigen-experienced datasets as somatic hypermutation continues during clonal expansion and so clonally related B cells may not express antibodies with identical CDR-H3.

Another issue with identifying clonal expansions is determining how the sequences should be compared. If the clone is identified at random and built around the first unique sequence identified, sequences at the "edges" of the clone may be placed differently in the clone than if, for example, the highest copy number sequence had been used as the starting point. Therefore, a better way of identifying clonal expansions is to take all potential members of a clone (e.g. all those with identical V gene and J gene usage) and to build a lineage tree. The branches of the tree are then "cut" and clones with a similar amount of mutation are considered part of the same clone (Hershberg & Luning Prak, 2015). The cut-off value chosen determines the level of similarity which defines a clonal expansion.

A common method of measuring CDR-H3 similarity is the Hamming distance (Hamming, 1950). The primary drawback of this method is that the CDR-H3 must be of the same length to be a clonal expansion. However, during sequencing, homopolymers common in the IGHD genes can lead to sequencing errors which can lead to clonally related CDR-H3 with apparently different lengths. Furthermore, SHM can sometimes introduce indels which could alter the length of the CDR-H3 (Wilson et al., 1998, Goossens et al., 1998). Therefore, an alternative distance measure, which does allow for indels, is the Levenshtein distance (Levenshtein, 1966).

It is the heavy chain (rather than the light chain) which is used for identifying clonal expansions as the diversity of the CDR-H3 means it would be very unlikely to find the same CDR-H3 twice by chance. The light chain is less diverse and public light chains can be found within multiple individuals (DeKosky et al., 2015). Therefore, although one would expect to find the same light chains within clonal expansions of paired heavy-light chain data, identifying clonally expanded light chains is not as reliable as with heavy chains. Also, because light chain rearrangement occurs in the bone marrow after some B cell proliferation, heavy chains from the same original B cell are found with different light chains (Wang et al., 2003).

1.8.3 Tools for Repertoire Analysis

IMGT High V-QUEST (Alamyar et al., 2012b, Alamyar et al., 2012a, Li et al., 2013, Giudicelli et al., 2015) has been available the longest and is the most popular. IMGT also maintains a number of immunoglobulin gene databases. There are a number of other tools available to identify VDJ genes. IgBLAST (Ye et al., 2013) is a modified version of the National Centre for Biotechnology Information (NCBI) BLAST tool (Altschul et al., 1997) which is aimed specifically at identifying immunoglobulin and T cell receptor genes, for example it is able to consider short (D gene) and long (V gene) matches simultaneously when analysing a rearranged variable region. IgDiscover (Corcoran et al., 2016) offers the capability to identify new V gene and alleles which are missing from current datasets, allowing the construction of personalised V gene datasets. JOINSOLVER (Souto-Carneiro et al., 2004) has a particular focus on analysing the CDR3. Other tools available include Ab-origin (Wang et al., 2008), VDJsolver (Ohm-Laursen et al., 2006), iHMMune-align (Gaëta et al., 2007), SoDA/SoDA2 (Volpe et al., 2006, Munshaw & Kepler, 2010) and VBASE2 (Retter et al., 2005). These methods use either sequence alignment or hidden Markov models to identify VDJ gene usage. IgSCUEAL (Frost et al., 2015) uses a phylogenetic approach to VDJ assignment.

Aside from assigning VDJ genes and identifying CDR3 regions, further analysis such as clonality analysis, somatic hypermutation frequency and CDR3 property analysis is possible. There are a number of suites available to do this, some are R packages, for example bcRep (Bischof & Ibrahim, 2016) and Change-O (Gupta et al., 2015). Others have a more user-friendly web interface, for example ARGalaxy (IJspeert et al., 2017), VDJServer (Christley et al., 2018) and BRepertoire (Margreitter et al., 2018).

1.9 Autoimmune Disease

Autoimmune disease occurs when the immune system attacks self-tissues, believing them to be pathogenic. Autoimmune diseases often involve autoreactive antibodies. Autoreactive antibodies can arise in a number of ways, primarily breakdown of tolerance, immunological ignorance (B cell did not encounter the autoantigen in the bone marrow) or as a result of an infection (molecular mimicry). The study of the B cell repertoire has allowed the identification of some common features of autoreactive antibodies such as frequent use of particular VDJ genes or a bias towards certain CDR3 properties (IGHV4-34 is very commonly associated with autoimmune disease). In some cases the B cell repertoire can be quite different in patients with an autoimmune disease compared to healthy controls (Dörner et al., 1999).

One of the most well-known autoimmune diseases is systemic lupus erythematosus (SLE), a disease in which multiple organs are attacked by the immune system. Autoantibodies are an important feature of SLE (over 100 autoantibody targets have been associated with the disease) however the autoantibodies most strongly associated with SLE are anti-nuclear antibodies (ANAs), which are present in over 95% of SLE patients (Bashford-Rogers et al., 2018) and are often present before the onset of clinical symptoms (Arbuckle et al., 2003). The autoantibodies in SLE are predominantly somatically mutated IgG and it is thought that they originate from autoreactive naïve B cells which are present in the periphery at a high frequency due to defective tolerance checkpoints (Yurasov et al., 2005, Yurasov et al., 2006). IGHV4-34 usage is very strongly associated with SLE (Tipton et al., 2015) and this gene is associated with a multitude of autoimmune targets, for example dsDNA, rheumatoid factors and I/i self-antigen on red blood cells (Potter et al., 2002). There is also a higher frequency of charged amino acids (such as arginine) in the CDR3 of B cells isolated from SLE patients (Bashford-Rogers et al., 2018). DNA is an intracellular antigen and therefore usually resides in an immune-privileged site. However, intracellular antigens such as DNA can become accessible in times of stress, disease or trauma.

Another example of autoimmune disease caused by antigen in an immune-privileged site is sympathetic ophthalmia. Sympathetic ophthalmia is an ocular autoimmune disease which occurs due to the release of immune-privileged antigens as a result of trauma to the eye (Margo & Harman, 2016).

Multiple sclerosis (MS) is thought to be an autoimmune disease which may be caused by molecular mimicry. MS has also been linked to infection such as Epstein-Barr virus (EBV), human herpes virus-6 (HHV-6), varicella zoster virus (VZV) and *Chlamydia pneumoniae* (Weber et al., 2011). A characteristic feature of MS is the presence of antibodies in the cerebral spinal

fluid and symptoms are caused by the gradual demyelination of the myelin sheath of neurones in the brain and spinal cord. However, identifying specific self-antigens associated with MS has proved challenging. Nevertheless, it has been shown that there is an increased frequency of IGHV1 and IGHV4 families (particularly IGHV1-69, IGHV4-34 and IGHV4-39) and also IGHD2, IGHD3 and IGHJ4 families in the brain lesions of MS patients (Baranzini et al., 1999), indicating that autoantibodies may be involved in demyelination.

Another well-known autoimmune disease is rheumatoid arthritis (RA). RA is the inflammation of the synovial membrane due to the activation of plasmablasts expressing autoantibodies against targets such as rheumatoid factor and citrullinated proteins (Tan et al., 2014). Dominant clones isolated from RA patient synovial fluid were enriched for IGHV4 usage (particularly IGHV4-34) and long CDR-H3 (Doorenspleet et al., 2014).

In addition to immunological ignorance and molecular mimicry, a further cause of autoreactivity is thought to be polyspecificity. It is thought that polyreactive antibodies bind multiple epitopes and are therefore likely to be reactive with a self-antigen. Despite this apparent danger, polyreactive antibodies are a common feature of a healthy immune antibody repertoire.

1.10 Polyreactivity

The traditional "lock and key" hypothesis of antibody binding suggested that antibodies have very rigid antigen-binding sites which would only bind a very particular complementary epitope. However, it was discovered that some antibodies were capable of binding multiple, structurally unrelated antigens. These antibodies are termed "polyreactive".

The mechanism of polyreactivity has been debated for some time and there is still no clear physicochemical or structural explanation or general rules for how polyreactive antibodies differ from monoreactive antibodies. One hypothesis is that a polyreactive antibody exists in multiple configurations/isomers before it even encounters antigen (Ma et al., 2002). Crystallographic studies have provided evidence of this (James et al., 2003). Another hypothesis is that the antigen-binding site is flexible and changes shape to fit various antigens ("induced fit" hypothesis) (Rini et al., 1992). This has been backed up by crystallographic studies which show, for example, that an anti-p24 (HIV1) monoclonal antibody binds a number of unrelated peptides, using different conformations and residues for binding (Keitel et al., 1997, Kramer et al., 1997). There is a hypothesis that some polyreactive antibodies have a hydrophobic "sticky" variable region due to aromatic residues (Padlan, 1994, Droupadi, 1994), however others have disputed this, arguing that polyreactive antibodies bind to structurally different antigens using different specific hydrogen bonds (James & Tawfik, 2003). Polyreactive antibodies tend to bind antigen with a lower affinity than monoreactive antibodies, they are often of IgM isotype and their variable regions tend to be closer to the germline configuration (Notkins, 2004). Animal studies suggest that polyreactive antibodies have a shorter half-life than monoreactive antibodies as they are cleared from the circulation much faster (Sigounas et al., 1994).

Polyreactivity is common in the natural antibody repertoire. Natural antibodies are antibodies that are secreted by B cells which have not been stimulated by cognate antigen. Natural antibodies are thought to be a first line of defence against an invading pathogen. The polyreactivity of natural antibodies allows them to bind new pathogens (albeit with a lower affinity than an antigen-matured antibody) and activate the complement pathway while a more specific humoral response is being mounted (Zhou et al., 2007). Furthermore, natural antibodies bind apoptotic cells to aid complement fixation and phagocytosis by macrophages (Gunti & Notkins, 2015). There is a high incidence of B cells expressing polyreactive antibodies in the thymus suggests that they may have a second role in tolerance, bringing self-antigen into the thymus from the periphery to help develop/maintain tolerance (Zhou & Notkins, 2004). Polyreactive antibodies appear to be an important part of the normal adult repertoire

and are the predominant type in newborns' repertoires (Chen et al., 1998). Polyreactive natural IgA antibodies may coat the microbiota in the gut (Bunker et al., 2017).

Polyreactivity can be beneficial in that it increases the diversity of the antibody beyond that achieved by V(D)J recombination, N-nucleotide addition and heavy-light chain pairing. However, the drawback of increased diversity is that it can result in autoreactivity. Some studies have found that polyreactive antibodies are prevalent in the bone marrow, but are selected out of the repertoire at the tolerance checkpoints (Wardemann et al., 2003). Antibody polyreactivity has been linked to autoreactivity as polyreactive antibodies have been found to be more prevalent in people with rheumatoid arthritis (Samuels et al., 2005) and SLE (Yurasov et al., 2005). However, a different study found the converse, suggesting that polyreactive antibodies were equally or less common in the repertoire of those with Sjögren's syndrome, rheumatoid arthritis and SLE (Chen et al., 1998). The studies used different methods to identify polyreactive antibodies; the former two used ELISA while the latter used FACS. This may have contributed to the differing results, however it does add to the conundrum of whether polyreactivity is advantageous or dangerous. There is plenty of evidence to suggest that polyreactivity is an important feature of natural antibodies, however they have also been found to be frequently autoreactive. Perhaps polyreactive antibodies are both advantageous and dangerous; a necessary evil to broaden the antibody repertoire and act as a first line of defence against infection, but with the risk of autoreactivity.

Furthermore, our understanding of exactly what polyreactivity is and how to define it remains unclear. The most accepted definition of polyreactivity is that it is the specific binding by an antibody of multiple structurally unrelated antigens. However, this is difficult to measure without solving the crystal structure of an antibody in complex with multiple antigens. Some polyreactivity may be down to molecular mimicry (although it is argued that this is not true polyreactivity as the same epitope is being bound by the antibody), whereas other polyreactivity may be down to antibody "stickiness" where hydrophobic patches interact with multiple antigens. But is such "stickiness" truly polyreactivity? Many antibodies may show some interaction with multiple antigens at a high enough concentration or under particular circumstances, but the interaction may be weak. It is argued by some that antibodies should only be described as polyreactive if they bind structurally different antigens with physiologically relevant affinities (Dimitrov et al., 2013). Another issue is how to identify a monoreactive antibody. How many different antigens need to be tested before an antibody is classified as monoreactive? Nobody has yet been able to identify general physicochemical or structural properties which can reliably differentiate polyreactive and monoreactive antibodies.

1.11 Antibody Response against Ebola Virus

1.11.1 Ebola Virus Disease

Ebola virus disease (more commonly referred to as Ebola) is a viral haemorrhagic fever caused by members of the genus *Ebolavirus*, which contains five species: *Zaire ebolavirus* (Ebola virus, EBOV), *Sudan ebolavirus* (Sudan virus, SUDV), *Taï Forest ebolavirus* (Taï Forest virus, TAFV), *Reston ebolavirus* (Reston virus, RESTV) and *Bundibugyo ebolavirus* (Bundibugyo virus, BDBV). The genetic diversity between *Ebolavirus* species is 25-35% (Grard et al., 2011). The *Ebolavirus* genus is part of the *Filoviridae* family, which includes two other genera: *Marburgvirus* and *Cuevavirus*. *Marburgvirus* contains two viruses of the species *Marburg marburgvirus*: Marburg virus (MARV) and Ravn virus (RAVV), both of which cause a severe haemorrhagic fever in humans similar to Ebola. *Cuevavirus* contains one species: *Lloviu cuevavirus* (Lloviu virus, LLOV) which is not thought to be pathogenic to humans. The phylogenetic relationships of the *Filoviridae* species are shown in Figure 1-12.

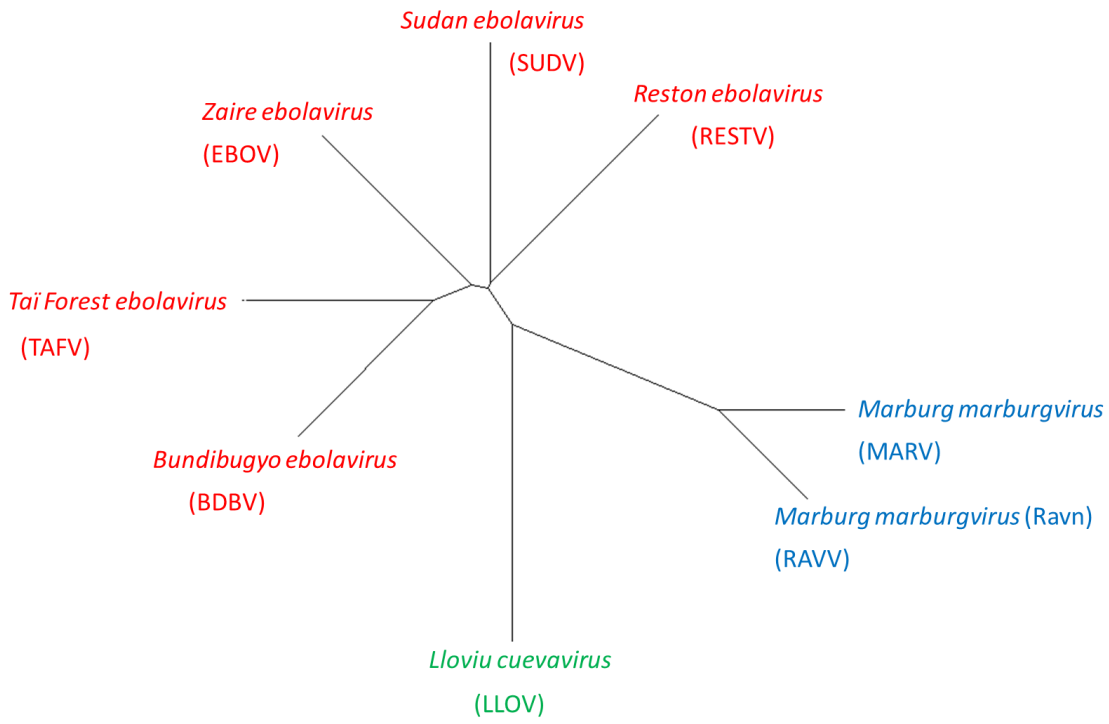


Figure 1-12: *Filoviridae* phylogenetic tree. The *Filoviridae* family contains three genera: *Ebolavirus* (red), *Marburgvirus* (blue) and *Cuevavirus* (green). Phylogenetic relationships were calculated by aligning complete viral genome sequences and displayed as an unrooted tree (Boc et al., 2012).

Symptoms of Ebola can include fever, headache, vomiting, diarrhoea, impaired kidney and liver functions, internal and external bleeding and, in some cases, death. Ebola spreads by direct human-to-human contact with infected individuals. The case fatality rate ranges from 25% to 90% depending on the outbreak and the *Ebolavirus* species responsible. The species most

frequently the cause of Ebola outbreaks, and also associated with the highest mortality, is *Zaire ebolavirus* (EBOV). EBOV was the cause of the 2014-2016 West African Ebola outbreak in which 28,610 people were infected and 11,308 people died in Sierra Leone, Liberia and Guinea (WHO, 2018a).

EBOV is coated with membrane-bound glycoprotein (GP) which is used for host cell entry by binding to cell surface molecules such as lectin-type molecules. The virion enters the host cell by a combination of macropinocytosis and endocytosis. Within the endosome, cathepsins proteolytically cleave the GP and the virus fuses with the endosome membrane, releasing viral ribonucleoprotein into the host cell cytoplasm where the viral genome is transcribed and viral replication occurs. Viral GP is expressed on the surface of infected host cells and new viral genomes are transported to the surface where they bud off, releasing new virus (Baseler et al., 2017).

Membrane-bound GP is a major target for antibodies. Therefore, EBOV employs a number of immune evasion tactics to hide membrane-bound GP from the immune system. Membrane-bound GP contains a very heavily glycosylated glycan cap and mucin-like domain (MLD) which shield the more conserved regions of the GP (Francica et al., 2010, Lennemann et al., 2014). Additionally, large amounts of secreted GP (sGP) (a truncated form of membrane-bound GP) is released extracellularly. It is thought that sGP acts as a decoy, binding GP-specific antibodies to prevent them from binding membrane-bound GP and neutralising the virus (Ito et al., 2001).

The main cell types infected by EBOV are dendritic cells, monocytes and macrophages and infected cells travel to the lymph nodes, via the lymphatic system, where the virus replicates and is disseminated. The viral load in the blood increases exponentially in the first three days following the onset of symptoms, becoming widely disseminated. This indicates that the virus overcomes innate immune defences. EBOV is able to evade both the innate and the adaptive immune responses using a number of mechanisms. Viral protein (VP) 24 and VP35 both antagonise the interferon response and VP35 also inhibits the upregulation of MHC class I, MHC class II and various costimulatory molecules, impeding the CD8 and CD4 T cell response and thus the entire adaptive immune response (Baseler et al., 2017).

Despite the immunosuppressive nature of Ebola, a strong antibody and T cell response is mounted during Ebola virus disease and antigen stimulation appears to continue for over a month after recovery as activated CD8 T cells persist for this length of time (McElroy et al., 2015). Much of the pathology associated with Ebola is thought to be due to the host immune system's response (e.g. release of cytokines).

It is unclear exactly when a patient has cleared the infection as viral RNA has been found to persist in other areas, such as urine and sweat, even after it no longer appears in the blood (Kreuels et al., 2014).

1.11.2 Viral Glycoproteins

Ebola viruses are enveloped, non-segmented negative-sense ssRNA viruses. The genome is approximately 19 kb and consists of seven genes (Figure 1-13A), all of which are monocistronic except the fourth gene in the genome (*GP*), which encodes three forms of glycoprotein (GP): envelope GP (GP), secreted GP (sGP), and small secreted GP (ssGP) (see Figure 1-13A). These three proteins are expressed from *GP* via transcriptional editing. The primary product is sGP and is translated from ORF 0. Approximately 71% of *GP* transcripts encode sGP. The secondary product is envelope GP. GP is expressed from ORF +1 after transcriptional editing where, during the transcription of seven uridines in the genomic RNA, an additional adenosine is added, causing a frameshift. Approximately 24% of *GP* transcripts encode GP. The tertiary product is ssGP. ssGP is the result of two adenosine insertions causing a frameshift to ORF +2. Approximately 5% of transcripts encode ssGP. sGP and ssGP are both secreted whereas envelope GP has a transmembrane domain and is therefore membrane-bound (Mehedi et al., 2011).

Envelope GP is first translated as a single peptide which is then cleaved with furin during post-translational modifications into GP1 and GP2 subunits. These subunits come together to form heterodimers held together by disulphide bonds and three heterodimers join to form the membrane-bound GP trimer (Jeffers et al., 2002) (Figure 1-13B). The more conserved regions of GP required for host cell entry and interactions in the endosome sit near the base of the GP or within the structure and are protected by the heavily glycosylated glycan cap and mucin-like domain (MLD). The glycan cap contains four N-linked glycosylation sites and the mucin-like domain contains five N-linked and 12-17 O-linked glycosylation sites (Lee & Saphire, 2009).

Marburg virus GP is only translated in a single reading frame and therefore expresses only membrane-bound GP. Like EBOV GP, Marburg GP is a trimer composed of three GP1-GP2 dimers (Jeffers et al., 2002).

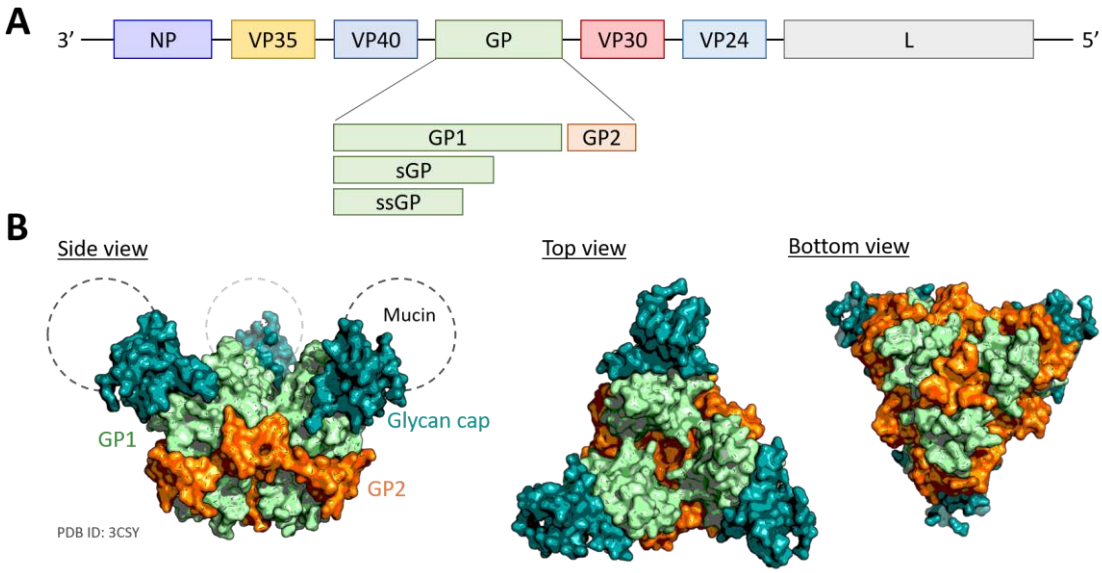


Figure 1-13: EBOV genome and structure of EBOV GP. **A)** *GP* is the fourth gene in the EBOV genome and encodes GP, sGP and ssGP; their expression is controlled by transcriptional editing. **B)** GP is a membrane-bound trimer composed of GP1-GP2 heterodimers. GP1 is shown in green, with the glycan cap highlighted in teal. GP2 is shown in orange. The location of the mucin-like domains are shown by grey circles in the side view diagram. PDB ID: 3CSY.

1.11.3 Antibody Response against Ebola Virus

The immune response to Ebola virus disease is still not fully understood. However, it is known that patients who have high anti-Ebola antibody titres early in Ebola infection are more likely to survive, whereas low or absent antibody titres are associated with fatalities (Ksiazek et al., 1999, Baize et al., 1999). This indicates that the antibody response is important in fighting Ebola virus disease.

Ebola survivor serology shows that the IgG response to Ebola is stronger than the IgM response; the IgM response declines after early infection (by day 80), whereas the IgG response persists (MacNeil et al., 2011). However both isotypes are capable of neutralising the virus and IgA is also raised against GP (Khurana et al., 2016). By using cell sorting, it has been shown that even three months after primary infection, approximately 3% of the IgG+ B cells in a patient's repertoire can still be EBOV GP-specific, which is similar to the frequency of antigen-specific antibodies seen in chronic HIV patients (Bornholdt et al., 2016b). Furthermore, circulating anti-EBOV GP antibodies are maintained for more than 10 years after infection (Misasi et al., 2016).

When investigating the composition of the B cell repertoire in response to Ebola, it has been found that the anti-GP B cell repertoire is very diverse (unlike anti-HIV and anti-influenza repertoires which show lots of clonal expansion). The repertoire is slightly skewed towards IgK chain usage and longer CDR-H3, but with no bias in V(D)J gene usage (Bornholdt et al., 2016b).

The most promising anti-Ebola vaccine is VSV-ZEBOV, which has just passed clinical trials. The vaccine raises an antibody response against Ebola (Zaire) glycoprotein (GP).

Vaccines are extremely important in the fight against Ebola, but so are post-exposure treatments as it can sometimes be difficult to get vaccines to everyone and sometimes vaccines do not confer complete, lifelong protection. Monoclonal antibody therapies are an important line of investigation for post-exposure treatments.

1.11.4 Anti-Ebola Therapeutic Antibodies

GP is the primary target of the antibody response against Ebola and membrane-bound GP trimers are the focus of efforts to find new antibody-based therapies. Antibodies have been discovered which bind a diverse range of antigenic sites on GP, for example the outward-facing glycan cap and mucin-like domain (MLD), the GP1-GP2 interface at the base of the structure and the ribosome-binding site (RBS) in a protected site within the GP (Figure 1-13B).

When investigating the potential of antibodies as therapeutic candidates, three stages of testing are initially employed:

1. Can the antibody bind to the GP *in vitro*? (e.g. ELISA, Biacore, Octet etc.)
2. Can the antibody neutralise the virus *in vitro*? (neutralisation assay)
3. Can the antibody protect animals *in vivo* from viral challenge?

To be a useful anti-GP antibody, an antibody must fulfil the first criterion of binding the GP structure. However, the antibody does not necessarily have to fulfil both criteria 2 and 3 to have therapeutic value. Some antibodies show neutralisation in *in vitro* neutralisation assays, but are not efficacious *in vivo*. Conversely, some antibodies may not neutralise an antibody *in vitro* but are still somehow effective *in vivo* or may be useful as part of a cocktail of monoclonal antibodies which bind several antigenic sites (Duehr et al., 2017).

Antibodies directed against different sites in the GP trimer have different *in vitro* neutralisation and *in vivo* protective efficacies. It is thought that antibodies which bind the GP1-GP2 interface at the base can neutralise the virus by preventing GP2 structural changes required for viral entry into host cells (Murin et al., 2014). Indeed, the human antibody KZ52 binds the GP1-GP2 interface and neutralises EBOV *in vitro* (Maruyama et al., 1999) and is also protective *in vivo* in guinea pigs (Parren et al., 2002). However, KZ52 is not protective *in vivo* in non-human primates (Oswald et al., 2007) indicating that the efficacy of an antibody relies on multiple factors, not just whether it binds the right epitope. Nevertheless, the GP1-GP2 interface is a good therapeutic target and other antibodies which bind in this region are effective in non-

human primates and humans. The best known examples of this are 2G4 and 4G7 which are part of the monoclonal antibody cocktails ZMab and ZMapp (Figure 1-14). Monoclonal antibody cocktails confer post-exposure protection and both ZMab and ZMapp were used to treat humans in the 2014 Ebola outbreak on compassionate grounds.

Figure 1-14: Binding sites of anti-EBOV GP monoclonal antibodies and composition of three well known anti-EBOV monoclonal antibody cocktails. ZMapp is the most efficacious cocktail and uses the best antibodies from MB-003 and ZMab. Image modified from González-González et al. (2017).

ZMab is composed of three mouse-derived anti-EBOV GP antibodies (1H3, 2G4 and 4G7). The cocktail provides 100% protection in macaques when administered within 24 hours of exposure and 50% protection when administered within 48 hours of exposure (Qiu et al., 2012). ZMapp also contains 2G4 and 4G7, however they are mixed with another mouse-derived antibody: 13C6 (Qiu et al., 2011, Wilson et al., 2000). ZMapp is more effective than ZMab as it shows 100% efficacy in non-human primates when administered up to five days post-infection (Qiu et al., 2014). The difference between the two cocktails is whether they use 1H3 (ZMab) or 13C6 (ZMapp). Both antibodies bind the glycan cap, indicating that the slight difference in binding site substantially affects efficacy of the cocktail.

Despite their clear value in monoclonal antibody cocktails, antibodies which bind in the glycan cap or MLC are rarely protective on their own or in the absence of complement (Murin et al., 2014). However, there are exceptions to this and protective antibodies have been identified which bind the glycan cap. For example, two human monoclonal antibodies (mAB100 and mAB114) used in combination protect non-human primates from lethal challenge with EBOV (Misasi et al., 2016). mAB100 bound at the GP1-GP2 base and mAB114 bound at the glycan cap. Surprisingly, monotherapy with mAB114 alone was sufficient to provide protection to macaques against lethal EBOV challenge (Misasi et al., 2016). Furthermore, broadly-neutralising antibodies have been isolated from the blood of human BDBV survivors which bind the glycan cap of multiple *Ebolavirus* species (Flyak et al., 2016). The identification of broadly-neutralising glycan cap-binders is particularly surprising because it is thought that the

main purpose of the heavily glycosylated glycan cap and MLD is to protect the more conserved GP epitopes (such as the RBS) from recognition by antibodies.

Due to the difficulty in creating anti-Ebola antibody therapies and the substantial threat from lesser-studied *Ebolavirus* species such as BDBV or RAVV or newly emerging Filoviruses (LLOV was discovered in Spain as recently as 2002) (Burk et al., 2016), cross-neutralising antibodies are particularly important. However they are difficult to find, particularly amongst all species in the *Filoviridae* family due to low GP sequence identity: EBOV and SUDV (the most divergent *Ebolavirus* species) have 56% GP sequence identity, whereas EBOV and MARV GP have just 30% sequence identity (Keck et al., 2016).

The most conserved GP region within the *Filoviridae* is the RBS and GP2 and these are often the sites of broadly-neutralising antibody binding. MR72 is a human antibody isolated from a MARV survivor which can neutralise VSVs expressing cleaved GP from all known Filoviruses. It does this by binding the highly conserved RBS (Bornholdt et al., 2016a). However, MR72 is unable to bind uncleaved GP expressed on the vesicular stomatitis virus (VSV) surface. This is probably because the RBS is well protected on the viral surface by the heavily glycosylated glycan cap and MLD. Therefore, any therapeutic anti-RBS antibodies may have to be targeted to the endosome where the GP is cleaved by proteases and the RBS is revealed. However, an RBS-binding antibody (FVM04) has been discovered which, when 4G7 in ZMapp was replaced by FVM04, the anti-EBOV efficacy of the ZMapp was retained but the cocktail also protected guinea pigs against SUDV (Howell et al., 2016). This indicates that RBS-binding antibodies may have a role in antibody cocktails to increase the breadth of protection.

EBOV-520 is a human IgG4 antibody which binds the base of GP. EBOV-520 shows *in vitro* neutralising capacity against EBOV, SUDV and BDBV and some *in vivo* protective capacity against these species in various animal models (Gilchuk et al., 2018). Furthermore, three antibodies isolated from BDBV human survivors (BDBV223, BDBV317 and BDBV340) bind to the bottom of the GP in the HR2 domain, close to the viral membrane and well below the GP1-GP2 interface. All three antibodies bind GP from EBOV, BDBV and SUDV. They also neutralise EBOV and BDBV *in vitro* and one of them neutralises RESTV (however they failed to neutralise SUDV). The antibodies also showed some protection against EBOV and BDBV in *in vivo* animal models (Flyak et al., 2018).

Cross-reactive antibodies between EBOV and SUDV have also been found in Macaques which bind at conserved epitopes in the GP1 core, glycan cap and fusion domain of GP2 (Keck et al., 2016).

As cross-protective antibodies are difficult to find, an alternative solution may be the use of bi-specific antibodies. For example, a bispecific antibody has been created which protects mice from EBOV and SUDV (Frei et al., 2016).

Post-exposure antibody therapies look to be a very promising treatment option against Ebola and we already have the efficacious monoclonal antibody cocktails ZMab and ZMapp. However, the greatest weakness of these treatments is that they are not cross-protective against multiple Filoviruses. Although many anti-GP antibodies have been characterised, the antibody, or mixture of antibodies, which will provide efficacious protection against multiple Filoviruses remains elusive. Antibodies isolated from human survivors are potentially very useful as they do not need to be humanised in order to be made safe for human use. However, the limited number of people who have been infected with Ebola makes this difficult. Furthermore, humans tend to be infected with just one strain, and so their antibodies are likely to be very strain-specific. Hybridoma and display technologies are therefore useful in the search for new cross-protective antibodies, which are essential in the development of monoclonal antibody therapies which could be used against any Filovirus outbreak.

1.12 Therapeutic Antibodies and their Discovery

Monoclonal antibody therapies are particularly useful for treating chronic diseases such as cancers and autoimmune diseases and some of the most famous include adalimumab (Humira), bevacizumab (Avastin), trastuzumab (Herceptin) and rituximab (Rituxan) which treat rheumatoid arthritis, colorectal cancer, breast cancer and non-Hodgkin's lymphoma respectively. Monoclonal antibody therapies for infectious diseases are less common as, to be most effective, they need to be administered as quickly as possible after infection. However, an anti-HIV antibody, ibalizumabuiyk (Trogarzo) has recently been approved and, as mentioned in the previous section, anti-Ebola monoclonal antibody cocktails have proved useful in the 2014 West Africa outbreak.

Monoclonal antibodies (and other types of antigen-specific molecules) have huge therapeutic potential and their discovery is a large part of many pharmaceutical companies' business. The two primary methods of isolating potentially therapeutic antibodies with the desired specificity is hybridoma and *in vitro* display technologies.

1.12.1 Hybridoma

Hybridoma is a method of generating antibodies using the immune response of an animal. The mouse (or other mammal) is challenged with the antigen of interest and it raises antibodies against the antigen. B cells are then harvested from the spleen of the animal and these are fused with an immortal B cell cancer cell line, creating the hybridoma cell line. The hybridoma cells are cultured from a starting point of one cell and each clonal population is screened for binding to the antigen of interest.

The two primary issues with hybridoma are firstly, if the antibodies are to be used in humans, the animal antibodies must be humanised. In addition to expressing the antibody using a human Fc, the Fab region must also be humanised by, for example, expressing the mouse CDR sequences on a human FR framework. Humanisation can be problematic because the process can sometimes lead to loss of specificity. Although most therapeutic antibodies are either human in origin or have been humanised, there are a small number of licensed murine and chimeric antibodies (Irani et al., 2015). The second issue with hybridoma is that animals cannot be used to raise antibodies against certain antigens, for example if the antigen is toxic to the animal.

1.12.2 Display Technologies

An alternative to hybridoma is display methods. These are *in vitro* methods of selecting for antibodies of interest in which a library is constructed of the immunoglobulin variable region sequences from the target species (often human). The variable regions are then displayed *in vitro* in a way that allows the phenotype (antibody variable region) to be linked to the genotype sequence which encoded it. The variable regions are screened for binding the target antigen *in vitro* (known as "panning selection") and the nucleotide sequences of the binders are identified. The variable region sequences can then be expressed as complete antibodies for further investigation.

The advantages of display methods over hybridoma are that animals are not used in the selection and the libraries can be used to find antibodies against almost any antigen. Furthermore, the synthetic libraries can be built from human repertoires, meaning that humanisation is not required. These synthetic libraries can be extremely large and can even be tailored to suit specific needs.

However, a drawback of synthetic libraries is that, because they are designed by humans, they can be vulnerable to researchers' preconceptions about what will make an effective antibody. For example, some libraries may rely on a limited range of V(D)J genes, or may introduce non-random diversity only to very specific regions. It has been shown that synthetic antibodies rely more heavily on the CDR-H3 for interactions with antigen than any other loop, making the antigen binding sites less complex than naturally occurring antibodies (Burkovitz & Ofran, 2016). Naturally occurring antibodies almost always use residues from at least three CDRs to make important contacts with antigen (Robin et al., 2014).

A further drawback of synthetic libraries is that heavy and light chains are almost always linked at random and so heavy-light chain pairs are not native. This may make it more difficult to find antigen-specific variable regions.

Variable regions are sometimes expressed as a whole Fab (scFab) but more commonly they are expressed as a single chain variable fragment (scFv). One reason for this is that Fab regions tend to be more toxic to *E. coli* and are expressed at a lower level. An scFv consists of the heavy and light chain variable regions of an antibody where, instead of being held together by the constant region, the variable regions are tethered together by a flexible linker peptide (Bird et al., 1988) (see Figure 1-15).

Linker regions can be 10-25 amino acids long and tend to incorporate hydrophilic amino acids. The most common flexible linker is the serine-glycine linker $(GGGGS)_3$, however other linkers,

such as ones incorporating part of the CH1/LH elbow regions (He et al., 1995, Tang et al., 2012), are sometimes used. It has been found that linker sequences composed of elbow regions bind equally well as those with serine-glycine linkers, however scFv are expressed in greater quantities by *E. coli* if the VL region precedes the VH region rather than vice versa (Anand et al., 1991).

There are multiple variations on the monomeric scFv. For example, by using shorter linker sequences, scFv can form diabodies or triabodies which allows for the possibility of bispecific or trispecific molecules.

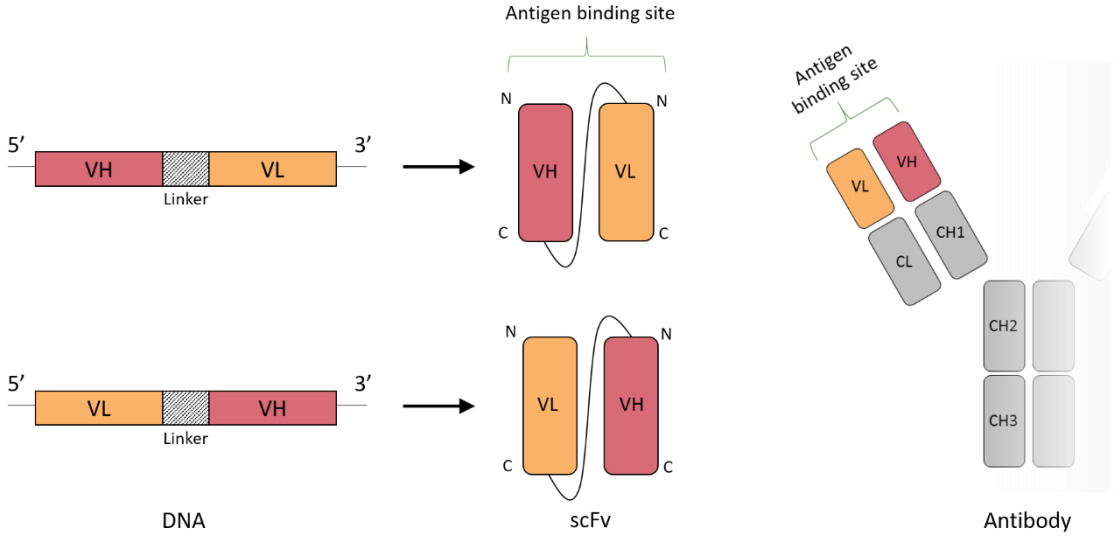


Figure 1-15: Single chain variable fragment (scFv) structure. Heavy and light chain antibody variable regions are joined by overlap extension PCR to form, when translated, an antigen binding site linked by a linker sequence instead of the CH1/CL region of a complete antibody. The VH sequence can precede the VL sequence or vice versa.

Despite the flexibility of the scFv format, it is usually important for a variable region to be expressed in a full antibody if it is to be use for therapeutic purposes. Most licensed therapeutic antibodies are full-length antibodies (predominantly IgG1 due to its prevalence in the blood and potent effector functions) (Irani et al., 2015) and it has been shown that the Fc region can be important for the protective activity of the antibody (Abboud et al., 2010, Bournazos et al., 2014). Although scFv libraries are usually built from human repertoires, and so humanisation is not required, binding specificity is not always maintained due to slight conformational changes.

The three most common display technologies for isolating antigen-specific scFv are phage display, ribosome display and yeast display. However, some less common techniques such as mRNA display and bacterial display are sometimes used.

1.12.2.1 Phage Display

Phage display is the most commonly used *in vitro* display technology. It involves the use of phage (usually M13 phage). The library is built by inserting the DNA encoding the scFv upstream of the gene encoding the geneIII terminal protein in the phagemid vector. Three to five copies of the scFv are expressed at the terminal end of the phage, attached to the geneIII protein (Lee et al., 2007) (Figure 1-16A).

To isolate antigen-specific scFv, the phage library is panned against the antigen of interest. The phage which do not bind are washed away and the phage which do bind are removed from the plate using trypsin protease which cleaves the c-Myc tag. The phage then infect TG1 *E. coli* via the F pili and helper phage induce the production and release of M13 phage from the *E. coli*. The rescued phage can then either be used in further panning selections or phage from individual *E. coli* colonies can be tested for antigen specificity using an ELISA (Lee et al., 2007) (Figure 1-16B).

The benefits of phage display are that the libraries are relatively large with an estimated functional size of 1 X 10^6 to 1 X 10^{11} (Chao et al., 2006). However the drawbacks are that as the scFv are expressed in *E. coli*, they are not subject to post-translational modifications such as glycosylation and the screening procedure is slow as it is done in a 96-well format.

Adalimumab (Humira), an anti-TNF mAb, was discovered using phage display (Lee et al., 2007).

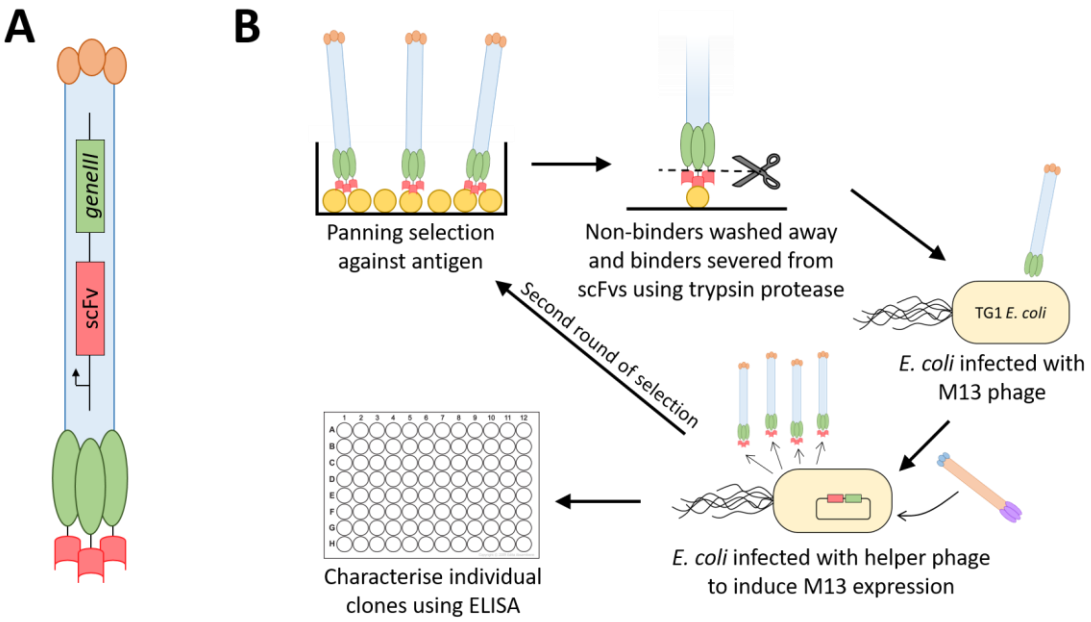


Figure 1-16: Phage display schematic. (**A**) The scFv library is inserted upstream of the geneIII expressing gene in M13 phage. (**B**) The M13 phage library is panned against the antigen. Antigen-binding M13 phage are used to infect TG1 *E. coli*. Helper phage cause the release of the M13 phage which can then be used either in another round of selection or in an ELISA. The diagram was based on the description of phage display set out in Lee et al. (2007).

1.12.2.2 Ribosome Display

Unlike a phage display library, which is stored in phage, a ribosome display library is maintained as DNA (either linear or in plasmids). The library is created by PCR and it is important that upstream of each variable region sequence there is a promoter (e.g. T7) and a ribosome binding site and at the 3' end it is very important that there is no stop codon. So that the library can be transformed into plasmids, restriction sites are also added at either end of the coding sequence. NcoI (5'-CCATGG-3') is a good restriction site for the 5' end as the ATG can also act as the start codon (Zahnd et al., 2007).

Before the panning selection, the DNA library is transcribed and translated *in vitro*. When the ribosome gets to the 3' end of the mRNA, as there is no stop codon, the ribosome cannot detach. The scFv and the mRNA are therefore linked by the ribosome (Figure 1-17A). A "spacer" nucleotide sequence is present downstream of the scFv-encoding region to sit in the ribosomal tunnel, allowing the scFv to protrude from the ribosome and fold correctly (Zahnd et al., 2007). A spacer is sometimes a non-functional sequence or could be part of the heavy or light chain constant region.

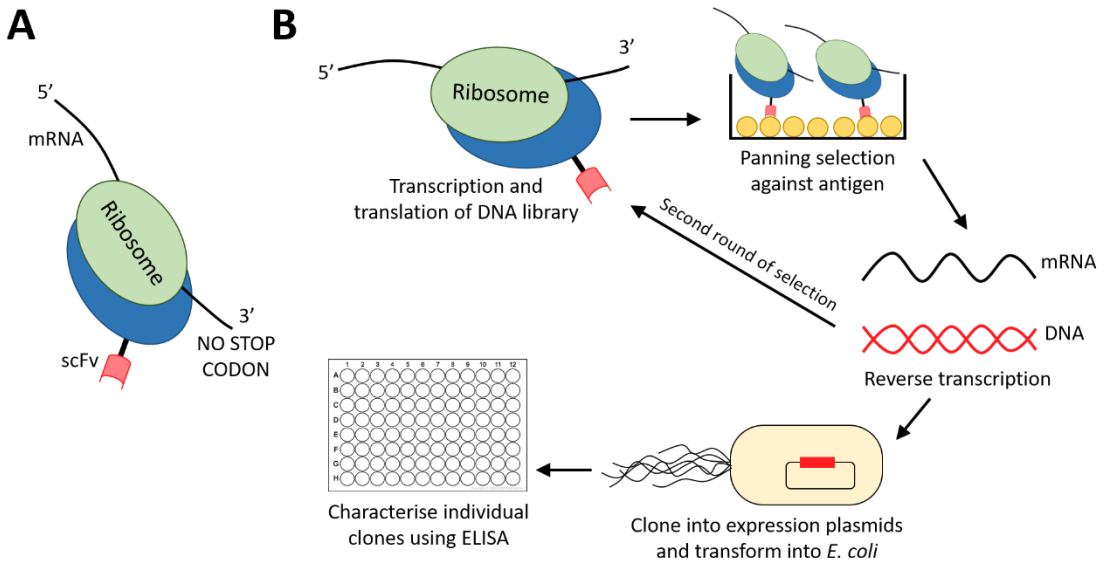


Figure 1-17: Ribosome display schematic. (**A**) When the ribosome display library is transcribed and translated, the ribosome gets stuck at the 3' end of the mRNA (no stop codon) and the scFv is joined to the coding mRNA. (**B**) The translated library is incubated on the antigen of interest. The non-binders are washed away and the mRNA from the binders is reverse transcribed and PCR amplified. The output library can either be used in a second round of panning selection or it can be inserted into expression plasmids and the scFv can be expressed in *E. coli* and characterised using ELISA.

The scFv-mRNA-ribosome complexes are panned against the antigen of interest. The non-binders are washed away and the binders are reverse-transcribed and amplified using PCR. The rescued library can then either be transcribed/translated and used again in further panning

selections or it can be inserted into an expression plasmid, transformed into *E. coli* and the specificity of scFv from individual *E. coli* colonies can be tested using ELISA.

The benefits of ribosome display are that the libraries are extremely large with an estimated functional diversity of 1×10^{11} to 1×10^{13} (Chao et al., 2006). However, like phage display, as the scFv are expressed in *E. coli* they do not get post-translational modifications. Also, like phage display, the screening is slow as it is also done in a 96-well format.

Also there is a large amount of PCR involved in building ribosome display libraries and in rescuing libraries after panning selections. Using such a large amount of PCR carries an increased risk of introducing mutations into the original immunoglobulin sequence. However, this can sometimes be turned into an advantage by using a low-fidelity DNA polymerase between rounds of panning selection to introduce random mutations and therefore create directed evolution where advantageous mutations are promoted with every round of selection. Higher-affinity scFv can therefore be created.

1.12.2.3 Yeast Display

The yeast display scFv library is displayed on the surface of each yeast cell at a frequency of 1×10^{4} t 1×10^{5} copies per cell. Each scFv is c-Myc-tagged and the scFv are fused to the yeast agglutinin protein (Aga2p). The successful expression of scFv by yeast can be detected by binding fluorescent tags to the c-Myc tag expressed on the scFv. Unlike phage display and ribosome display, in yeast display the scFv are expressed on a eukaryotic cell. Therefore to perform a panning selection, the yeast library is mixed with biotinylated antigen, which can be detected with fluorescently-labelled avidin. The yeast library is then sorted in a cell sorter to isolate the antigen-binders (Chao et al., 2006).

One of the benefits of yeast display is that the scFv are expressed in a eukaryotic system and they therefore undergo post-translational modifications (unlike phage and ribosome display libraries which are expressed in prokaryotes). The size of the libraries is much smaller than phage and ribosome display libraries with an estimated functional diversity of 1×10^{7} to 1×10^{9} (Chao et al., 2006). However, the libraries can be screened for binders very efficiently using cell sorters and if multiple fluorescent colours are used, multi-specific binders can be screened. This means that millions of scFv could be screened in just a few hours, as opposed to screening lots of 96-well plates using ELISA. However, yeast have a slower doubling time than *E. coli*, meaning that the selection process takes longer.

1.13 Project Aims

The overarching aim of this thesis is to better understand the relationship between antibody variable region properties and their specificity or potential role in the immune response. The primary method we used to investigate this was the high-throughput sequencing of various B cell repertoires. Further computational and wet lab experiments were then used to complement this work. The aims of the three results chapters are summarised below.

Chapter 3: Tolerance, Autoreactivity and Polyreactivity

A consequence of the huge diversity of the B cell repertoire is that autoreactive and/or polyreactive antibodies can arise. Such antibodies can cause autoimmune disease and are problematic in the pharmaceutical industry as it is imperative that biologics are highly specific and not autoreactive. The aims of this first chapter were to use the B cell Development dataset (the first high-throughput sequencing dataset created from human bone marrow and peripheral blood B cells) to identify heavy chain properties which may be selected against at the tolerance checkpoints and therefore may be associated with autoreactivity. We aimed then to test these predictions of autoreactivity/polyreactivity using immunocytochemistry and ELISA. It was our hope that using a high-throughput sequencing approach may allow us to identify common, predictive characteristics of autoreactive/polyreactive antibodies and that this information would be a useful timesaver in the search for potentially therapeutic antibodies.

Chapter 4: Light Chain Selection and Isotype Properties

In the literature (and also in Chapter 3) the focus of investigations into antibody specificity is often on the heavy chain as experiments have shown that the heavy chain usually makes more contacts with antigen than the light chain. However, light chains are important for antibody diversity and swapping light chains is the mechanism by which receptor editing rescues autoreactive heavy chains during central tolerance. Therefore the aim of this chapter was to investigate the properties of antibody light chains. We investigated two aspects of antibody light chains. Firstly, using the B cell Development dataset, we investigated which light chain properties appeared to be under selection as B cells matured through tolerance and therefore may be associated with autoreactivity/polyreactivity. Secondly, we compared the physicochemical and structural properties of kappa and lambda light chains to investigate whether the two classes of light chain had similar or different properties and therefore whether they may play different roles in the immune response. Further, we investigated the

properties of the heavy chains with which kappa and lambda light chains are paired to determine if we could identify any evidence of bias in heavy-light chain pairing.

Chapter 5: The Antibody Response against Ebola

The final aim of this thesis was, instead of considering antibody properties in the context of non-specific autoreactive and polyreactive responses, to investigate the B cell repertoire in response to a specific infection: Ebola virus disease. Ebola is a devastating disease which puts a huge toll on patients and has a very high mortality rate. Serology shows that a strong antibody response is linked to survival and studies looking at Ebola-specific antibodies and B cells have shown that the major target of the antibody response is the glycoprotein (GP) which coats the virus and it uses for host cell entry. Monoclonal antibody cocktails such as ZMapp are a promising post-exposure therapy, however a major drawback of those available so far is that they do not offer protection against multiple filoviruses. Therefore the aims of this chapter were to use high-throughput sequencing to attempt to investigate the properties of the B cell repertoires of recent Ebola survivors and how they differ from negative controls who have never suffered the disease. Additionally, we aimed to isolate new cross-reactive anti-Ebola GP antibodies using ribosome display.

Chapter 2

Materials and Methods

Chapter 2: Materials and Methods

2.1 High-Throughput Sequencing: B cell Development Dataset[1]

The B cell Development dataset is a human B cell repertoire high-throughput sequencing dataset constructed from immunoglobulin mRNA from sorted B cells at four different stages of development (pre-B, immature, transitional and naïve). The B cells were isolated from the bone marrow and peripheral blood of healthy adult human donors. Immunoglobulin mRNA was reverse-transcribed using random hexamers and the cDNA was amplified using V gene-specific and constant region-specific primers. Multiplex identifier (MID) sequences were added using PCR to identify the donors in the data analysis. Heavy and light chain sequences were not paired. See Figure 2-1 for a schematic.

The B cell Development dataset is used in Chapter 3 (Tolerance, Autoreactivity and Polyreactivity) and Chapter 4 (Light Chain Selection and Isotype Properties).

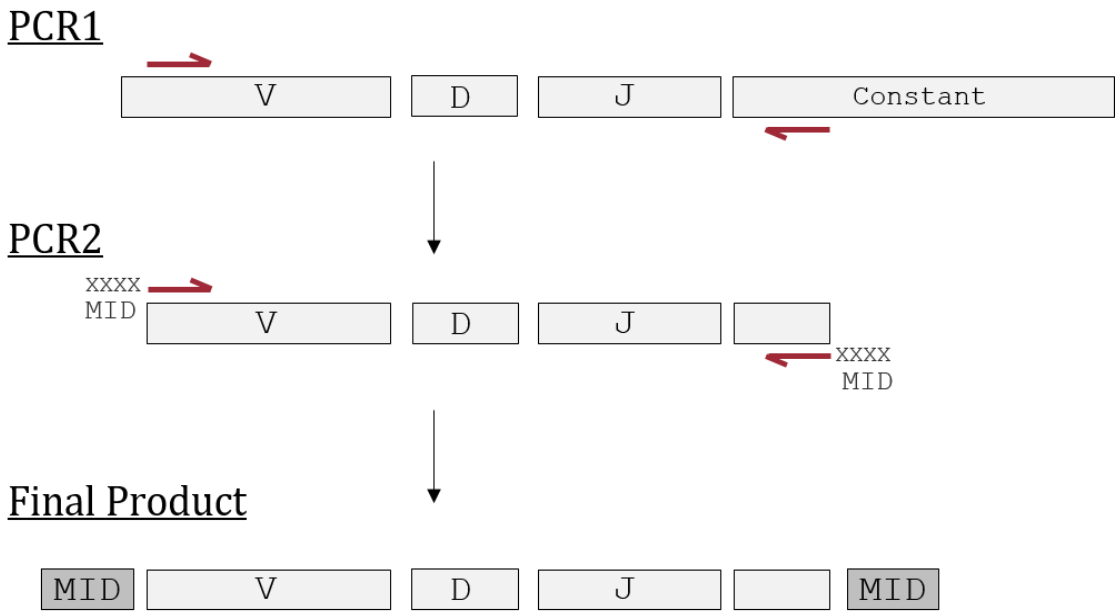


Figure 2-1: PCR1 and PCR2 used to construct the B cell Development dataset. Random hexamers were used to reverse transcribe mRNA from isolated B cells. In PCR1, the forward primers (6X IgH, 6X IgK, 5X IgL) bind slightly downstream of the 5' end of the V gene. The reverse primers bind close to the beginning of the IgM CH1 region or IgK/IgL constant regions. In PCR2, the primers bind in the same places but include a multiplex identifier (MID) which identifies the sample. Primers are represented in pink.

[1] The B cell Development dataset was constructed in 2012. All work described in section 2.1 was conducted by Bryan Wu (KCL), David Kipling (Cardiff University), Joselli Silva-O'Hare (KCL), Victoria Martin (KCL) and Deborah Dunn-Walters (University of Surrey). This method has been published in Wu et al. (2015) and Townsend et al. (2016).

2.1.1 Sample Collection

19 healthy adults undergoing hip replacement surgery at Guy's Hospital, London, United Kingdom were recruited. The head of the femur was collected as the source of bone marrow and a sample of peripheral blood was also collected (Research Ethics Committee: 14/LO/1221). The patients ranged in age from 24 years to 86 years and had no known autoimmune disease or recent infections.

2.1.2 Isolation of Lymphocytes from Bone Marrow

The bone marrow matrix was scraped out of the femur head and residual cells were washed out of the cavity using BM isolation buffer (10 mM EDTA, 2% (v/v) HI-FCS, 1X PBS, pH 7.4) and transferred to a 50 ml graduated centrifuge tube. The volume was made up to 35 ml with BM isolation buffer. The cell suspension was then filtered through a 100 µm cell strainer. The cell suspension was layered onto 15 ml room temperature Ficoll (GE Healthcare) and the bone marrow mononuclear cells (BMMCs) were isolated by centrifugation at 400 g for 35 min (no break). The BMMC layer was removed, re-suspended in BM isolation buffer, and centrifuged at 300 g for 10 min. The pelleted BMMCs were re-suspended in 1 ml RPMI suspension buffer (10% (v/v) HI-FCS, RPMI 1640) and centrifuged in a bench-top centrifuge at 2X 1000 for 5 min. The pelleted BMMCs were re-suspended in 1 ml RPMI suspension buffer and counted. The BMMCs were frozen in HI-FCS, 10% (v/v) DMSO at -80^{o}C (Mr. Frosty Freezing Container, Thermo Scientific) and then transferred to liquid nitrogen for long-term storage.

2.1.3 Isolation of Lymphocytes from Peripheral Blood

10 ml of peripheral blood was mixed with 20 ml RPMI suspension buffer (section 2.1.2) and the 30 ml blood solution was layered onto 15 ml Ficoll (GE Healthcare) in a Leucosep tube (Greiner). This was centrifuged at 448 g for 20 min (no break). The PBMC layer was transferred into a clean 50 ml graduated centrifuge tube, re-suspended in 10 ml RPMI suspension buffer and centrifuged at 275 g for 10 min. The pelleted PBMCs were re-suspended in 10 ml RPMI suspension buffer and centrifuged at 275 g for 10 min. The pelleted PBMCs were re-suspended in RPMI suspension buffer, counted, and frozen as described in section 2.1.2.

2.1.4 B cell Isolation and Sorting

Bone marrow CD19+ B cells were enriched to >98% using anti-human CD19 MicroBeads (Miltenyi Biotec) following the manufacturer's instructions. The CD19+ enriched B cell suspension was blocked using blocking buffer (RPMI, 10% (v/v) normal mouse serum). The B cells were stained using the antibodies detailed in Table 2-2. 1000-10,000 large pre-B cells (IgK-IgL-IgM+CD38+) and 1000-10,000 immature B cells (IgK+/IgL+IgM+IgD-CD38+CD27-CD10+)

per donor were then sorted directly into PCR tubes containing 36 µl SLyRT buffer (Table 2-1) using the FACS Aria (BD Biosciences). Sorting strategy is shown in FACS plots in Figure 2-2A.

Peripheral blood B cells were stained using the antibodies detailed in Table 2-3. 1000-10,000 transitional B cells (IgD+CD27−CD10+) and 1000-10,000 naïve B cells (IgD+CD27−CD10−) per donor were sorted directly into PCR tubes containing 36 µl SLyRT buffer (Table 2-1) using the FACS Aria (BD Biosciences). Sorting strategy is shown in FACS plots in Figure 2-2B.

Table 2-1: SLyRT buffer reagents. Scale up as necessary. The final reaction volume will be 40 µl after cDNA, SuperScript III Reverse Transcriptase and primers are added

Reagent	1X Reaction	Final Reaction Concentration
First-Strand RT Buffer (5X)[2]	8 µl	1X
Random hexamers pd$(N)_6$ (50 ng/µl)	12 µl	15 ng/µl
5% (v/v) Triton X-100 in H_2O	1 µl	0.13% (v/v)
RiboSafe RNase Inhibitor (40 U/µl)[3]	2.5 µl	2.5 U/µl
DTT (0.1 M)	4.5 µl	11.25 mM
dNTP mix (10 mM each)[4]	2 µl	500 µM
H_2O	6 µl	-

Table 2-2: Fluorescently-labelled antibodies used to sort BMMCs.

Antibody	Clone	Manufacturer
PE anti-human Ig light chain lambda	MHL-38	Biolegend
APC anti-human Ig light chain kappa	MHK-49	Biolegend
PE/Cy7 anti-human CD38	HIT2	Biolegend
PerCP/Cy5.5 anti-human IgD	IA6-2	Biolegend
Pacific Blue anti-human IgM	MHM-88	Biolegend
APC/Cy7 anti-human CD10	HI10a	Biolegend
CD27-FITC	M-T271	Miltenyi Biotec

Table 2-3: Fluorescently-labelled antibodies used to sort PBMCs.

Antibody	Clone	Manufacturer
APC anti-human CD19	HIB19	BD Bioscience
PerCP/Cy5.5 anti-human IgD	IA6-2	Biolegend
APC/Cy7 anti-human CD10	HI10a	Biolegend
CD27-FITC	M-T271	Miltenyi Biotec

[2] First-Strand RT Buffer (5X): 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$

[3] RiboSafe RNase Inhibitor (BIO-6502, Bioline)

[4] dNTP mix: 10 mM each of dATP, dTTP, dCTP and dGTP (Promega)

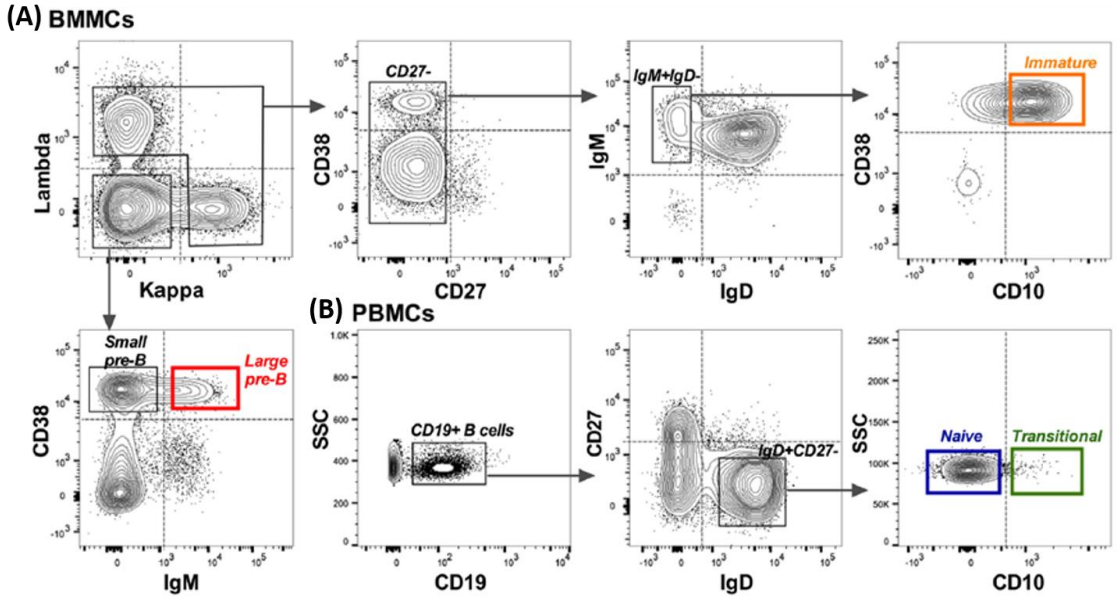


Figure 2-2: FACS plot showing B cell sorting strategy of (**A**) BMMCs and (**B**) PBMCs to isolate pre-B, immature, transitional and naïve B cells. Image modified from Martin et al. (2016)[5]

2.1.5 Reverse Transcription of Immunoglobulin mRNA

The PCR tubes containing sorted B cells were mixed by vortexing and centrifuged at 3000 rpm for 2 min. 4 µl of 25 U/µl SuperScript III Reverse Transcriptase (18080093, Invitrogen) was added to each tube. Reverse transcription was performed using the conditions shown in Table 2-4.

Table 2-4: Reverse transcription thermal cycle.

Thermal Cycle Step	Temperature	Time
Denaturation	42°C	10 min
Annealing	25°C	10 min
RT Extension	55°C	60 min
RT Termination	72°C	15 min

To remove cellular contaminants for future steps, the PCR tubes were centrifuged at 13,000 rpm for 5 min and 35 µl of supernatant was transferred to new PCR tubes (leaving cell debris behind). 70 µl of H_2O was added to the 35 µl supernatant (mixed by pipetting). This results in a 1:3 dilution of cDNA.

2.1.6 Amplification of cDNA (PCR1)

To amplify the immunoglobulin sequences from the cDNA, V gene-specific forward primers and constant region-specific reverse primers were used. Constant region-specific reverse primers (IgM, IgK and IgL) were used in separate reactions. Multiple V gene-specific forward

[5] Figure 2-2 is a modified version of a figure created by Victoria Martin (KCL) and Deborah Dunn-Walters (University of Surrey) and published in Martin et al. (2016).

primers were used in combination with each constant region reverse primer in a multiplex PCR. See Table A-1 (page 261) for a list of the primers used to amplify IgM sequences, IgK sequences and IgL sequences.

A PCR1 master mix (Table 2-5) was made for each class (IgH/IgK/IgL) using the appropriate primers (Table A-1). Each PCR1 master mix was aliquoted into eight PCR tubes (22 µl per tube) per donor. This means that if a donor had IgM, IgK and IgL sequences amplified, there would be 24 PCR reactions for that donor.

3 µl of 1:3 diluted cDNA (section 2.1.5) from each donor was added to each of their respective PCR tubes. The conditions for PCR1 were as described in Table 2-5.

Table 2-5: HTS PCR1 master mix reagents and thermal cycle.

Reagent	Quantity (X1 reaction)	Quantity (X110 reactions)
Phusion GC Buffer (5X)[6]	5 µl	550 µl
1:40 Phusion Polymerase (0.125 U/µl)[6]	5 µl	550 µl
dNTP mix (20 mM each)[7]	0.25 µl	27.5 µl
Primer mix[8]	1.25 µl	137.5 µl
H_2O	10.5 µl	1155 µl
Total	22 µl	2420 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X15
Annealing	58°C	15 s	
Extension	72°C	30 s	
Final extension	72°C	10 min	X1
Hold	4°C	∞	

2.1.7 Addition of MIDs (PCR2)

In order to identify the donor from which each sequence originated, 10 bp multiplex identifier (MID) sequences were added to PCR1 products (section 2.1.6) using a second PCR (PCR2). Due to the complexity of the MID-labelled primers, the PCR2 master mix was made up without primers (Table 2-6) and the relevant primer mix was then added to each tube. A separate primer mix was used per isotype (V gene-specific and constant region specific) and per donor (MID labels). See Table A-2 (page 262) and Table A-3 (page 262) for primer sequences. Each primer mix contained 835 nM each forward primer and 5 µM of the reverse primer.

17 µl of PCR2 reaction mix (Table 2-6) was aliquoted into 16 PCR tubes per isotype/donor combination (i.e. 48 tubes per donor if IgM, IgK and IgL were being amplified). 1 µl of the

[6] Phusion High-Fidelity DNA Polymerase (M0530, New England Biolabs)
[7] dNTP mix: 20 mM each of dATP, dTTP, dCTP and dGTP (Promega)
[8] Primer mix: 5' end variable primers (835nM each) and 3' end constant primer (5 µM). See Table A-1 for sequences.

appropriate MID-labelled isotype-specific primer mix was added into each tube. 2 µl of relevant PCR1 product (section 2.1.6) was added into each set of 16 PCR tubes (i.e. two PCR2 reactions per PCR1 reaction). The conditions are shown in Table 2-6.

The MIDs used to identify each donor in the B cell Development dataset are shown in Table 2-7 (the same MID can be used more than once if the samples are sequenced on separate chips; 12 MIDs can be used per chip).

Table 2-6: HTS PCR2 master mix reagents and thermal cycle.

Reagent	Quantity (X1 reaction)	Quantity (X110 reactions)
Phusion GC Buffer (5X)[9]	4 µl	440 µl
1:20 Phusion Polymerase (0.25 U/µl)[9]	2 µl	220 µl
dNTP mix (20 mM each)[10]	0.2 µl	22 µl
H_2O	10.8 µl	1188 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X20
Annealing	58°C	15 s	
Extension	72°C	30 s	
Final extension	72°C	10 min	X1
Hold	4°C	∞	

Table 2-7: MIDs used in B cell Development dataset to identify donors and cell repertoires.

Patient ID	Age (years)	Gender	Chip ID	MID to identify patient and cell repertoire			
				Pre-B	Immature	Transitional	Naïve
122	24	Female	BW062012a	MID10	MID4	MID7	MID1
159	28	Female	BW062012a	MID11	MID5	MID8	MID2
138	41	Male	BW062012a	MID12	MID6	MID9	MID3
107	43	Male	HTS TD	-	-	MID3	MID9
120	44	Male	HTS TD	-	-	MID4	MID10
118	48	Male	BW062012c	MID10	MID4	MID7	MID1
146	50	Female	BW062012c	MID11	MID5	MID8	MID2
103	52	Female	BW062012c	MID12	MID6	MID9	MID3
128	52	Female	N, T, Imm cells	MID11	MID12	MID1	-
126	53	Male	N, T, Imm cells	MID3	MID4	MID5	MID6
141	65	Female	BW062012d	MID10	MID4	MID7	MID1
105	67	Female	N, T, Imm cells	MID7	MID8	MID9	MID10
119	68	Female	BW062012d	MID11	MID5	MID8	MID2
160	70	Male	BW062012d	MID12	MID6	MID9	MID3
111	71	Female	BW062012b	MID12	MID6	MID9	MID3
149	72	Male	BW062012b	MID11	MID5	MID8	MID2
132	76	Male	HTS TD	-	-	MID5	MID11
162	78	Male	HTS TD	-	-	MID6	MID12
140	86	Female	BW062012b	MID10	MID4	MID7	MID1

[9] Phusion High-Fidelity DNA Polymerase (M0530, New England Biolabs)
[10] dNTP mix: 20 mM each of dATP, dTTP, dCTP and dGTP (Promega)

2.1.8 Isolation and Purification of Variable Region Amplicons

Each set of 16 PCR2 products (i.e. each MID/Ig gene combination) was pooled (320 µl each). 64 µl of 6X Orange G loading dye (0.25 g Orange G (Sigma Aldrich), 30 ml glycerol, 70 ml H_2O) was added to each pooled PCR2 product. The mixtures were heated at 98°C for 5-10 min and then loaded onto a 1.5% agarose gel (TAE buffer) stained with GelStar Nucleic Acid Gel Stain (Lonza). The samples were run alongside Hyper Ladder IV ladder (Bioline) at 100 V for 60 min. Fragments of the correct size (Table 2-8) were excised from the gel and transferred to 15ml graduated centrifuge tubes.

Table 2-8: Approximate HTS PCR2 fragment sizes.

PCR2 Fragment	Approximate size
IgM	400 bp
IgK	500 bp
IgL	500 bp

DNA was purified from the gel using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's quick start protocol with the following modifications:

- Step 2: Incubate the gel slice in 5 ml Buffer QG until dissolved (expect about 6 ml total)
- Step 5: Pass the entire 6 ml gel + QG Buffer mixture through a single QIAquick column 750 µl at a time, saving the flow-through in a new 15 ml graduated centrifuge tube instead of discarding it. Repeat by passing the flow-through through the same QIAquick column a second time
- Step 9: Elute the DNA in 55 µl H_2O (incubate at room temperature for 3 min before centrifugation). Repeat this step by applying an additional 55 µl to the membrane, incubating for 3 min and eluting (final volume of eluted DNA is approximately 100 µl)

The concentration and quality of the eluted DNA was measured using the Qubit dsDNA HS Assay Kit (Q32851, Invitrogen) on a Qubit 2.0 Fluorometer.

2.1.9 Preparation of Library for Roche 454 Sequencing

The purified PCR2 products were mixed into a single sequencing sample containing 900 µl per MID. If a donor's heavy and light chains were sequenced, 300 µl of IgM, IgK and IgL were mixed. If just the donor's light chains were sequenced, 450 µl of IgK and IgL were mixed.

The library was concentrated using the Qiaquick PCR purification kit (Qiagen) according to manufacturer's instructions and was eluted in 110 µl H_2O (two lots of 55 µl as in section 2.1.8).

The concentrated library was quantified using Qubit dsDNA HS Assay Kit (Q32851, Invitrogen) on the Qubit 2.0 Fluorometer. For sequencing, the library needed to be a minimum of 5 µg DNA at a concentration of 100 ng/µl.

The library was sequenced across three chips (Table 2-7, page 79) on the Roche 454 GS FLX Titanium platform (LGC Genomics, UK).

2.2 High-Throughput Sequencing: Ebola Response Datasets

Unlike the B cell Development dataset (section 2.1), the Ebola Response datasets were built using mRNA from TRIzol-treated whole blood (the B cells could not be sorted for safety reasons) therefore B cell types could not be distinguished. The template switch method was used to add unique identifier (UID) sequences to each cDNA as it was reverse-transcribed, allowing better identification of clonal expansions. Forward primers bound the universal primer site (reducing primer bias) and reverse primers bound the constant region (allowing distinction between IgA, IgG and IgM classes and subclasses). PacBio identifier (PID) sequences were added using PCR to identify donors in the data analysis. Heavy and light chains were not paired. See Figure 2-3 for a schematic.

When the Ebola Response Dataset was built, the Roche 454 sequencing platform was no longer supported and the sequencing reads we produced for this dataset were longer than for the B cell Development dataset. Therefore PacBio sequencing was used instead. One drawback of PacBio sequencing is its relatively high error rate (Rhoads & Au, 2015). However, this platform does support long sequencing reads, making it a suitable choice for this experiment.

The Ebola Response Datasets were used in Chapter 5 (The Antibody Response against Ebola).

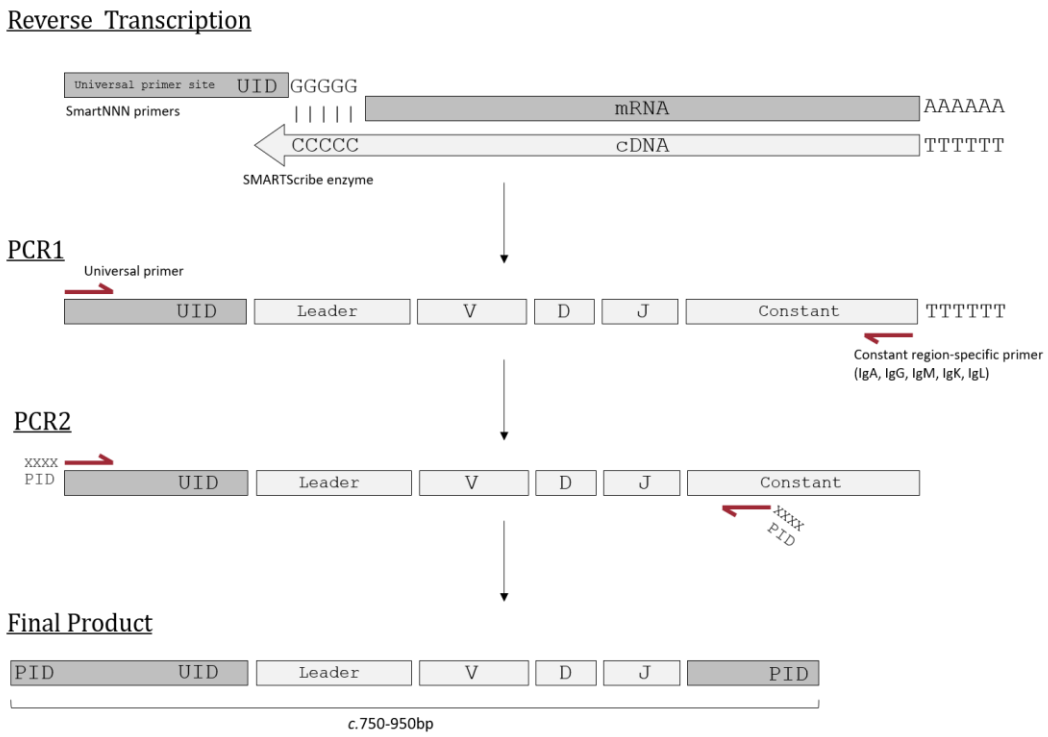


Figure 2-3: HTS template switch method used to generate the Ebola Response datasets. Amplification of immunoglobulin variable region sequences from mRNA for high-throughput sequencing (HTS). Unique identifier sequences (UID) were added to each mRNA using SmartNNN primers. PIDs were added using primers in PCR2. Once PIDs have been added, all sequences can be pooled for HTS on a single PacBio sequencing chip. Primers are represented in pink.

2.2.1 Sample Collection

2.2.1.1 Caucasian Aid Worker Survivor Dataset[11]

TRIzol-treated whole blood[12] from three Caucasian aid worker Ebola (EBOV) survivors from the 2014 West Africa outbreak. Blood was collected after recovery, approximately one month after infection. No serology was available for these samples. RNA from these donors was used to build a high-throughput sequencing dataset and also a ribosome display library (see section 2.14, page 118). (14/LO/1221, RES committee, Camberwell St Giles Substantial amendment #1, February 2015).

No blood samples from before the aid workers were infected were available. Therefore PBMCs from the whole blood of three age- and gender-matched donors (see Table 2-9) who had never been exposed to Ebola were used as negative controls. Blood from the negative control donors had been collected before (day 0) and 28 days after (day 28) receiving a yellow fever vaccine (14/LO/1221, RES committee, Camberwell St Giles).

Table 2-9: Caucasian Aid Worker Survivor Dataset donor information.

Group	Donor ID	Age	Gender	Time of Sample
Ebola Survivors	215	29	Male	In recovery
	216	34	Male	In recovery
	217	39	Female	In recovery
Yellow Fever Vaccine Controls	193	27	Male	Day 0 Day 28
	199	34	Male	Day 0 Day 28
	207	36	Female	Day 0 Day 28

[11] The Caucasian Aid Worker Survivor Dataset was built in 2016 by Joselli Silva O'Hare (KCL), Audrey Page (KCL) and Catherine Townsend (author)

[12] TRIzol-treated whole blood from three Caucasian aid worker Ebola survivors was generously shared by Alain Townsend (University of Oxford)

2.2.1.2 African Survivor Dataset[13]

mRNA from the whole blood of African survivors[14] of the 2014 West Africa Ebola (EBOV) outbreak was used to build the African Survivor Dataset. Blood had been taken at several time points post-recovery from Ebola virus disease and serology information on the IgG and IgM response was also provided (Figure 2-4). RNA from two time points from five donors was used to build the African Survivor Dataset. The time points used are shown in Figure 2-4. (14/LO/1221, RES committee, Camberwell St Giles Substantial amendment #1, February 2015).

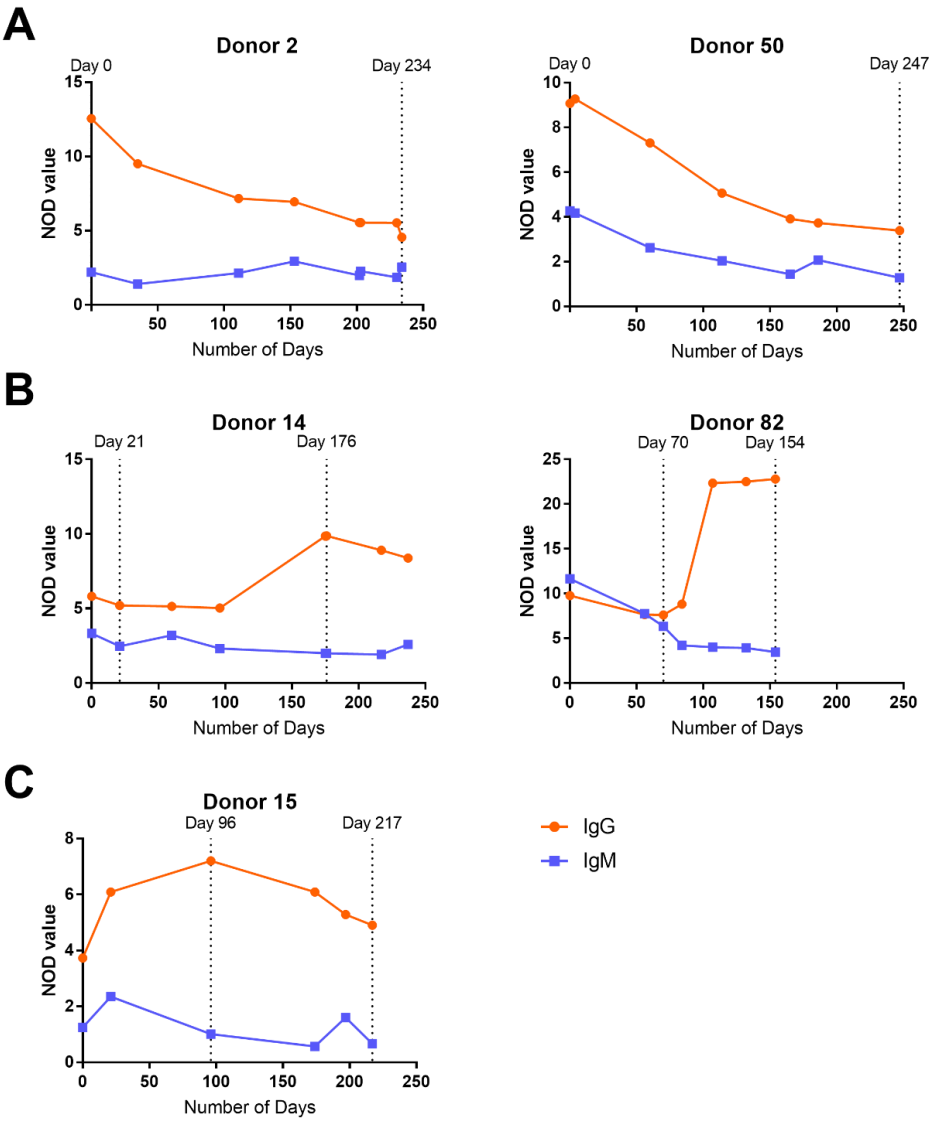


Figure 2-4: IgG and IgM serology from the five African Ebola survivors[15] used to make the African Survivor Dataset. The IgG serology of the five survivors showed different patterns over time. Note that time point zero is the first post-recovery time point on which blood was taken. (**A**) IgG started high and decreased over time. (**B**) IgG rose over time. (**C**) IgG was consistently high. The two time points used are shown by dotted lines. NOD = "normalised optical density".

[13] The African Survivor Dataset was built in 2017 by Audrey Page (KCL) and Alexander Stewart (University of Surrey)
[14] Purified total RNA from African Ebola survivors was generously shared by Callum Semple (University of Liverpool)
[15] The serology data graphed in Figure 2-4 was supplied by Callum Semple (University of Liverpool), the data was graphed by Audrey Page (KCL)

2.2.2 Isolation of Total RNA from Samples

2.2.2.1 Caucasian Aid Worker Ebola Survivors[16]

Blood samples were received as blood cone filters in TRIzol Reagent (15596, Invitrogen) (2.5 ml blood in 7.5 ml TRIzol Reagent). The samples were stored at -80°C until use.

The samples (2.5 ml blood in 7.5 ml TRIzol Reagent) were thawed and pipetted up and down several times. The samples were incubated at room temperature for 5 min. 2 ml of chloroform was added per 10 ml sample. The tubes were shaken vigorously for 15 s, then allowed to stand for 10 min at room temperature. The samples were centrifuged at 12,000 g for 15 min at 2-8°C. The colourless upper aqueous phase was transferred to a new tube.

5 ml of 100% isopropyl alcohol was added to each sample. The samples were mixed and left to stand for 10 min at room temperature. The samples were centrifuged again at 12,000 g for 15 min at 2-8°C. The RNA formed a pellet down the side of the tube. The supernatant was discarded and the RNA pellet washed with 10 ml 75% (v/v) ethanol. The sample was vortexed and centrifuged at 7,500 g for 5 min at 2-8°C (plus a further 5 min at 12,000 g if the RNA pellet was floating in the sample). The supernatant was discarded and the RNA pellet was left to dry for 5-10 min. 30 µl of H_2O was added to elute the RNA. To facilitate dissolving of the RNA, the sample was mixed by pipetting and incubated at 55-60°C for 10-15 min. The RNA was quantified.

2.2.2.2 Yellow Fever Vaccine Controls[17]

50 ml of blood was taken from healthy donors before receiving the yellow fever vaccine (day 0) and 28 days after (day 28) receiving the live attenuated yellow fever vaccine. Donors received the vaccine at several clinics around London, United Kingdom.

PBMCs were isolated by layering on Ficoll. Briefly, the blood sample was diluted in suspension media (RPMI-1640 (11875093, Gibco), 10% (v/v) FCS) in a 1:2 ratio. 35 ml of diluted blood was layered on top of 15 ml Ficoll-Paque Plus (17-1440-02, GE Healthcare). The samples were centrifuged at 400 g for 30 min (no break) at room temperature. PBMCs were collected using a Pasteur pipette and transferred into a clean 50 ml centrifuge tube. Suspension media was added and the PBMCs were centrifuged at 400 g for 10 min at 4°C. The supernatant was discarded and the PBMCs were re-suspend in 5 ml suspension media, then transferred to a clean 15 ml centrifuge tube. The PBMCs were counted using a haemocytometer. The PBMCs

[16] All of the work described in section 2.2.2.1 was conducted by Joselli Silva O'Hare (KCL)

[17] In section 2.2.2.2, the PBMC isolation from whole blood was conducted by Joselli Silva-O'Hare (KCL) and Victoria Martin (KCL). The total RNA isolation from the frozen PBMCs was conducted by Audrey Page (KCL) and Catherine Townsend (author)

were centrifuged at 400 g for 10 min at 4°C. The PBMCs were re-suspend in ice-cold freezing media (90% (v/v) FCS, 10% (v/v) DMSO). 1 ml of 1.5 X 10^6 cells/ml PBMCs were aliquoted into cryovials. The PBMCs were frozen overnight at -80°C in Mr Frosty (Nalgene) isopropanol containers then transferred to liquid nitrogen for long-term storage.

Two 1 ml PBMC aliquots were thawed per YF vaccine control sample in a 37°C waterbath until a small piece of ice remained. 2 ml pre-warmed media (RPMI, 10% (v/v) FCS) was added to 15 ml centrifuge tubes. The PBMCs were carefully added to the media using a pipette. The mixture was topped up to 10 ml using the pre-warmed media (RPMI, 10% (v/v) FCS). The samples were incubated on ice until cold and then centrifuged at 300 g for 10 min at 4°C. The supernatant was decanted and the cells were re-suspended in 1 ml of media (RPMI, 10% (v/v) FCS). Cells were counted by taking a sample to mix in a 1:1 ratio with trypan blue and using a haemocytometer. The 1 ml cell suspension was topped up to 10 ml with media (RPMI, 10% (v/v) FCS) and centrifuged at 300 g for 10 min at 4°C. The supernatant was discarded and the cells were re-suspended in 1 ml sterile PBS. Cells were stored on ice until RNA extraction.

RNA was extracted from PBMCs using the Quick-RNA MicroPrep kit (R1050, Zymo Research) following the manufacturer's protocol for cells in suspension with the following modifications and notes:

1. Cells were pelleted at 300 g for 5 min at 4°C, all the PBS was removed from the cells and the cells were re-suspended in 300 µl lysis buffer
2. The Sample Clearing step was performed
3. The samples were eluted in 10 µl H_2O (the H_2O was incubated on the column for 1 min prior to elution to maximise the amount of RNA eluted)

The RNA was stored at -70°C until use.

2.2.2.3 African Ebola Survivors

Samples were received as purified total RNA. Therefore samples could be reverse transcribed with no prior treatment.

2.2.3 Reverse Transcription of Immunoglobulin mRNA

Oligo(dT) primers were used to reverse transcribe all mRNA. To denature the RNA and anneal the Oligo(dT) primers Mix 1 was set up as shown in Table 2-10. The amount of RNA per sample must be identical, so all samples are scaled to the sample with the lowest RNA concentration.

To perform the reverse transcription using the SmartNNN template switch adapter primer, Mix 2 was set up as shown in Table 2-11.

Table 2-10**:** Ebola HTS Mix 1 reagents and thermal cycle.

Reagent	Quantity
Oligo(dT)$_{15}$ Primer (10 µM)[18]	1 µl
RNA	327 ng
H2O	To 4 µl
Total	4 µl

Thermal Cycle Step	Temperature	Time	Cycles
Denaturation	70°C	4 min	X1
Annealing	42°C	2 min	X1

Table 2-11: Ebola HTS Mix 2 reagents and thermal cycle. See Table A-4 for primer sequence.

Reagent	Quantity
5X First-Strand Buffer[19]	2 µl
DTT (20 mM)[19]	1 µl
SmartNNN primer (10 µM)	1 µl
dNTP Solution Mix (10 mM each)[20]	1 µl
SMARTScribe Reverse Transcriptase[19]	1 µl
Mix 1 (Table 2-10)	4 µl
Total	10 µl

Thermal Cycle Step	Temperature	Time	Cycles
Reverse Transcription	42°C	60 min	X1

The SmartNNN template switch adapter primer contains deoxyuridine. After cDNA synthesis, uracil-DNA glycosylase (UDG) was used to eliminate the SmartNNN primers to prevent the exchange of UIDs during PCR amplification.

1 µl of UDG (M0280L, New England Biolabs) was added per reaction tube and was heated at 37°C for 60 min. The UDG was deactivated by heating at 95°C for 10 min.

[18] Oligo(dT)$_{15}$ Primer (C1101, Promega)
[19] SMARTScribe™ Reverse Transcriptase (639537, Clontech)
[20] Deoxynucleotide (dNTP) Solution Mix (N0447L, New England Biolabs)

2.2.4 Amplification of Immunoglobulin Variable Regions

Immunoglobulin variable region were amplified from cDNA in PCR1 as shown in Table 2-12. Eight PCR1 reactions were run per sample.

Table 2-12: Ebola HTS PCR1 reagents and thermal cycle. See Table A-5 for primer sequences.

Heavy Chain Reaction Reagent	Quantity		Light Chain Reaction Reagent	Quantity	
	X1	X8		X1	X8
5X Q5 reaction buffer[21]	4 µl	32 µl	5X Q5 reaction buffer[21]	4 µl	32 µl
dNTP mix (10 mM each)[22]	0.4 µl	3.2 µl	dNTP mix (10 mM each)[22]	0.4 µl	3.2 µl
IGHA-1st primer (10 µM)	1 µl	8 µl	IGLC-1st primer (10 µM)	1 µl	8 µl
IGHG-1st primer (10 µM)	1 µl	8 µl	IGKC-1st primer (10 µM)	1 µl	8 µl
IGHM-1st primer (10 µM)	1 µl	8 µl			
Smart20 primer (10 µM)	1 µl	8 µl	Smart20 primer (10 µM)	1 µl	8 µl
Q5 enzyme (NEB)[21]	0.2 µl	1.6 µl	Q5 enzyme (NEB)[21]	0.2 µl	1.6 µl
H_2O	10.4 µl	83.2 µl	H_2O	12.4 µl	99.2 µl
cDNA (Table 2-11)	1 µl	8 µl	cDNA (Table 2-11)	1 µl	8 µl
Total	20 µl	160 µl	Total	20 µl	160 µl
Thermal Cycle Step	**Temperature**		**Time**	**Cycles**	
Initial denaturation	98°C		30 s	X1	
Denaturation	98°C		10 s	X21	
Annealing	65°C		20 s		
Extension	72°C		50 s		
Final extension	72°C		5 min	X1	
Hold	4°C		∞		

In the second PCR (PCR2), primers labelled with PIDs were used to amplify the PCR1 product. The addition of PIDs allowed the identification of the sample source in the bioinformatic analysis and allowed all samples to be pooled and sequenced on the same sequencing cells. 16 PCR2 reactions were run per sample (two reactions per PCR1 reaction). The reagents and conditions for PCR2 are shown in Table 2-13. See Table 2-14 and Table 2-15 for the PIDs used in each dataset.

[21] Q5 High-Fidelity DNA Polymerase, M0491L, New England Biolabs
[22] Deoxynucleotide (dNTP) Solution Mix, N0447L, New England Biolabs

Table 2-13: Ebola HTS PCR2 reagents and thermal cycle. See Table A-6 for primer sequences.

Heavy Chain Reaction Reagent	Quantity		Light Chain Reaction Reagent	Quantity	
	X1	X16		X1	X16
5X Q5 reaction buffer[23]	4 µl	64 µl	5X Q5 reaction buffer[23]	4 µl	64 µl
dNTP mix (10 mM each)[24]	0.4 µl	6.4 µl	dNTP mix (10 mM each)[24]	0.4 µl	6.4 µl
Step_1-Barcode primer (10 µM)	1 µl	16 µl	Step_1-Barcode primer (10 µM)	1 µl	16 µl
IGHA/IGHG/IGHM-Barcode primer (10 µM)	1 µl	16 µl	IGKC/IGLC-Barcode primer (10 µM)	1 µl	16 µl
Q5 enzyme[23]	0.2 µl	3.2 µl	Q5 enzyme[23]	0.2 µl	3.2 µl
H_2O	12.4 µl	198.4 µl	H_2O	12.4 µl	198.4 µl
PCR1 DNA	1 µl	16 µl	PCR1 DNA	1 µl	16 µl
Total	20 µl	320 µl	Total	20 µl	320 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X12
Annealing	65°C	20 s	
Extension	72°C	50 s	
Final extension	72°C	5 min	X1
Hold	4°C	∞	

Table 2-14: PIDs used in Caucasian Aid Worker dataset. See Table A-7 for PID sequences.

Donor ID	Age (years)	Gender	Sample Type	PID
215	29	Male	EBOV Survivor	PID1
216	34	Male	EBOV Survivor	PID2
217	39	Female	EBOV Survivor	PID4
193	27	Male	Yellow Fever Control Day 0	PID17
			Yellow Fever Control Day 28	PID44
199	34	Male	Yellow Fever Control Day 0	PID9
			Yellow Fever Control Day 28	PID33
207	36	Female	Yellow Fever Control Day 0	PID22
			Yellow Fever Control Day 28	PID45

Table 2-15: PIDS used in African Survivor Dataset. See Table A-7 for PID sequences.

Group	Donor ID	Time point	PID
IgG starts high, decreases over time	Donor 2	Time 1	PID1
		Time 2	PID2
	Donor 50	Time 1	PID4
		Time 2	PID9
IgG rose over time	Donor 14	Time 1	PID17
		Time 2	PID33
	Donor 82	Time 1	PID22
		Time 2	PID23
IgG consistently high	Donor 15	Time 1	PID44
		Time 2	PID45

[23] Q5 High-Fidelity DNA Polymerase, M0491L, New England Biolabs
[24] Deoxynucleotide (dNTP) Solution Mix, N0447L, New England Biolabs

PCR2 amplification was checked by running 5 µl PCR2 product mixed with 1 µl 6X gel loading dye (B7024S, New England Biolabs) on a 1.2% (w/v) agarose gel at 110 V for 1 h. The agarose gel was stained with GelStar Nucleic Acid Gel Stain, 50535, Lonza. The expected band sizes are shown in Table 2-16.

Table 2-16: Ebola HTS expected PCR2 band sizes. Range of sizes is due to variation in CDR3 size.

Ig class	Expected band size
IgA	779-845 bp
IgG	756-831 bp
IgM	749-824 bp
IgK	793-868 bp
IgL	864-939 bp

The PCR2 samples were pooled by combining each of the 16 PCR2 reactions per patient/isotype and purifying the PCR product using the Wizard SV Gel and PCR Clean-Up System (A9281, Promega) with the following alterations to the manufacturer's protocol:

- After the PCR product had been spun through the membrane once (Binding of DNA, step 3), the flow-through was collected, applied to the column again and the step was repeated to ensure maximum binding of PCR product.
- After washing the column with 500 µl washing buffer (Washing, step 5), the column was transferred to a new collection tube and the column was re-centrifuged (Washing, step 6) for 5 min.
- The DNA was eluted from the column in two elution steps (Elution, step 8), first in 30 µl H_2O, then a further 10 µl H_2O was applied to the column and the column was spun again (i.e. 40 µl in total).

Patient samples were processed one at a time to minimise the risk of cross-contamination.

2.2.5 Isolating Amplicons of the Right Size

The five bands (IgA, IgG, IgM, IgK, IgL) were isolated from each donor sample using the Pippin Prep (SAGE Science).

The pooled PCR products from section 2.2.4 were topped up to 30 µl with H_2O and each was mixed with 10 µl of loading solution with marker K (CDF1510, SAGE Science) and loaded into 1.5% dye-free agarose gels (one gel per donor). The Pippin Prep was set to collect bands within the range 500-1060 bp for every isotype.

The concentration of the DNA collected by the Pippin Prep was quantified using Qubit 3.0 (Invitrogen) and the Qubit dsDNA HS Assay Kit (Q32851, Invitrogen). All samples had to contain

at least 100 ng of DNA. If a sample did not have enough DNA, more PCR2 product was amplified and purified (page 89 onwards) for that sample.

2.2.6 Pooling Samples and Sequencing

Equal quantities of DNA for each isotype and donor were pooled. The mixture was purified using the Agencourt AMPure XP PCR Purification kit (A63880, Beckman Coulter). The pooled PCR product was aliquoted into 100 µl volumes and the protocol for 96-well format was followed. For maximum PCR product binding to magnetic beads, the optional 5 min incubation (step 3) was performed. The DNA was eluted in 40 µl H_2O per tube. The purified products were then pooled.

The Qubit 3.0 was used as described above to quantify the eluted DNA. There needed to be at least 500 ng of DNA at a concentration of at least 13 ng/µl.

A sample was the sent to the Earlham Institute (Norwich, United Kingdom) for high-throughput sequencing using single-molecule real-time (SMRT) sequencing on the PacBio RS II sequencing platform.

2.3 High-Throughput Sequencing: Ribosome Display Libraries

To investigate the diversity of the ribosome display input libraries and the selection that took place, the ribosome display input and output libraries (described in section 2.14, page 118) used to select for EBOV GP-binding scFv were sequenced using high-throughput sequencing.

2.3.1 Sample Collection

The starting material was the PCR3 product input library (section 2.14.4, page 123) and the PCR4 product output library (section 2.15.1, page 126). Donor 216 did not have enough PCR product. Therefore only PCR product from donor 215 and donor 217 was used to build the HTS library. Kappa and lambda PCR products were amplified separately.

2.3.2 Amplification of Input and Output scFv Sequences

The PCR3 input libraries were at a higher concentration than the PCR4 output libraries. Therefore, the PCR3 input DNA was diluted with dH_2O until the PCR3 and PCR4 products for each donor were at approximately equal concentrations. Four 8X PCR master mixes (Table 2-17) were set up per donor (kappa and lambda PCR3 and PCR4). The master mixes were aliquoted into 8X 20 µl PCR reactions and run using the conditions shown in Table 2-17. The PIDs used for each sample are shown in Table 2-18.

Table 2-17: Ribosome display HTS PCR reagents and thermal cycle conditions. Primer sequences are shown in Table A-8 (page 264).

Reagent	Quantity	
	X1	X8
5X Q5 reaction buffer[25]	4 µl	32 µl
dNTP mix (10 mM each)	0.4 µl	3.2 µl
RDT7AB_PID primer (10 µM)	1 µl	8 µl
NotK/NotL_PID primer (10 µM)	1 µl	8 µl
Q5 enzyme (NEB)[25]	0.2 µl	1.6 µl
H_2O	12 µl	96 µl
DMSO	0.4 µl	3.2 µl
PCR3/PCR4 Kappa[26]/Lambda[27] DNA	1 µl	8 µl
Total	20 µl	160 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X12
Annealing	67°C	30 s	
Extension	72°C	60 s	
Final extension	72°C	5 min	X1
Hold	4°C	∞	

[25] Q5 High-Fidelity DNA Polymerase, M0491L, New England Biolabs
[26] Donor 215: Kappa PCR3 input DNA = 15.3 ng/µl, kappa PCR4 output DNA = 13.7 ng/µl
Donor 217: Kappa PCR3 input DNA = 6.1 ng/µl, kappa PCR4 output DNA = 6.7 ng/µl
[27] Donor 215: Lambda PCR3 input DNA = 6.2 ng/µl, Lambda PCR4 output DNA = 6.9 ng/µl
Donor 217: Lambda PCR3 input DNA = 8.8 ng/µl, Lambda PCR4 output DNA = 8.2 ng/µl

Table 2-18: PID use for each ribosome display sample. PIDs attached to RDT7AB, NotK and NotL primers to label samples by donor and PCR3 input/PCR4 output. The same PIDs were used for kappa and lambda sequences.

Donor ID	Sample	Age (years)	Gender	Sample Type	PID
215	PCR3	29	Male	EBOV Survivor	PID1
215	PCR4	29	Male	EBOV Survivor	PID9
217	PCR3	39	Female	EBOV Survivor	PID4
217	PCR4	39	Female	EBOV Survivor	PID22

2.3.3 Isolating Amplicons of the Right Size

Each group of eight PCR reactions was pooled and the pooled PCR reactions were concentrated using the Wizard SV Gel and PCR Clean-Up System (A9281, Promega) as described in section 2.2.4 (page 88). PCR products of the correct size were then purified using the Pippin Prep (SAGE Science) using the protocol described in 2.2.5 (page 90). The Pippin Prep was set to collect bands within the range 800-1100 bp. The concentration of eluted DNA was confirmed using the Qubit 3.0 (Invitrogen) and the Qubit dsDNA HS Assay Kit (Q32851, Invitrogen).

2.3.4 Pooling Samples and Sequencing

All samples were pooled and purified using the Agencourt AMPure XP PCR Purification kit (A63880, Beckman Coulter) as described in section 2.2.6 (page 91) except the DNA was eluted in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0).

The sample was sequenced using PacBio Sequel sequencing at the Centre for Genomic Research (University of Liverpool).

2.4 High-Throughput Sequencing: Raw Data Clean-Up[28]

2.4.1 Removal of Biologically Implausible Sequences

Raw sequencing data is received in FASTA format. Firstly, biologically implausible sequences and sequences which contained PCR artefacts (e.g. sequences which were too short to be immunoglobulin sequences or the presence of internal or mismatched MIDs/PIDs) were removed. The MID/PID sequences were then removed from the nucleotide sequences and the sequences were renamed to contain the sample information which was associated with the MIDs/PIDs.

2.4.2 V(D)J Gene Usage and CDR3 Amino Acid Sequence Properties

The sequences were submitted to HighV-QUEST (Alamyar et al., 2012b) for biological annotation such as the identification of V(D)J genes and CDR3 amino acid sequences. Physicochemical properties of the CDR3 amino acid sequences were calculated using the R package Peptides (Osorio et al., 2015). The properties calculated were: GRAVY (Kyte & Doolittle, 1982), Boman (Boman, 2003), aliphatic index (Ikai, 1980), isoelectric point (pI) (Rice et al., 2000), instability (Guruprasad et al., 1990) and the ten Kidera factors (Kidera et al., 1985a, Kidera et al., 1985b) (see Table 2-19 for Kidera factor definitions). The frequency of certain classes of amino acids (Rice et al., 2000) in the CDR3 regions was also calculated (Table 2-20).

The CDR regions were defined using the IMGT numbering scheme (see Figure 1-6, page 33), which is the same for heavy and light chains. The CDR1 comprised residues 27-38, the CDR2 comprised residues 56-65, the CDR3 comprised residues 105-117.

Table 2-19: Kidera factor definitions as listed in Rackovsky &Scheraga (2011).

Kidera factor	Definition
Kidera 1	Helix/bend preference
Kidera 2	Side-chain size
Kidera 3	Extended structure preference
Kidera 4	Hydrophobicity
Kidera 5	Double-bend preference
Kidera 6	Partial specific volume
Kidera 7	Flat extended preference
Kidera 8	Occurrence in alpha region
Kidera 9	pK-C
Kidera 10	Surrounding hydrophobicity

[28] All of the work described in section 2.4 was conducted by David Kipling (Cardiff University) and Deborah Dunn-Walters (University of Surrey)

Table 2-20: Amino acid categories as defined in Rice et al. (2000).[29]

Category	Amino acids
Small	ABCDGNPSTV
Tiny	ACGST
Polar	DEHKNQRSTZ
Non-polar	ACFGILMPVWY
Aromatic	FHWY
Aliphatic	AILV
Charged	BDEHKRZ
Acidic	BDEZ
Basic	HKR

2.4.3 Clonotype Clustering

Clonotype clustering was carried out by splitting the datasets into V family subsets. The CDR3 nucleotide sequences were used to generate Levenshtein edit distance matrices of all possible pairwise comparisons. The matrices were then hierarchically clustered (complete linkage). The heavy chain dendrograms were cut at 0.18. The light chain dendrograms were cut at 0.05. The released branches constituted clones. The modal sequence of each clone was used as the representative of that clone. In most analyses, only the representative modal sequences were used to avoid large clones skewing the analysis.

In the Ebola Response Datasets, UIDs were used to differentiate clonal expansions and PCR amplification, therefore PCR amplifications (identified by duplicate UIDs) were removed from clonal expansions. In the B cell Development Dataset UIDs were not used therefore it was not possible to differentiate clonal expansions and PCR amplifications. However, as the B cell Development dataset is an antigen-naïve dataset, it should not contain clonal expansions. As the Ribosome Display Library was amplified from DNA and was already the product of multiple rounds of PCR, UIDs were not used. However, clonal expansions were still calculated in order to estimate diversity and therefore the efficiency of experimental selection.

[29] A = Alanine, B = Asparagine or Aspartic Acid, C = Cysteine, D = Aspartic acid, E = Glutamic acid, F = Phenylalanine, G = Glycine, H = Histidine, I = Isoleucine, K = Lysine, L = Leucine, M = Methionine, N = Asparagine, P = Proline, Q = Glutamine, R = Arginine, S = Serine, T = Threonine, V = Valine, W = Tryptophan, Y = Tyrosine, Z = Glutamine or Glutamic Acid

2.5 High-Throughput Sequencing: Data Analysis

2.5.1 VDJ Gene Usage

Pivot tables in Microsoft Excel 2010 were used to calculate the frequency with which heavy and light chain VDJ families and genes were used in each donor. Kappa and lambda sequences were analysed separately. Only a single, modal, sequence from each clonotype was used to calculate mean family/gene frequency (+/- 1 SD) to ensure the frequencies were not skewed by PCR amplification or clonal expansions.

To ensure the VDJ family/gene frequencies being compared were based on meaningful data, the "5X rule" was used to exclude any samples that contained too few observations. The 5X rule stipulates that if the sequence count of a sample was less than five times the number of families/genes being investigated, that sample was excluded from the analysis because the sequence count was too low to give a meaningful representation of gene frequency.

In the B cell development light chain dataset, a small number of light chain sequences were amplified from pre-B cells. During gene frequency analysis and CDR-L3 property analysis, pre-B sequences were grouped with the immature sequences and analysed collectively as immature B cells. These cells were likely to have been transitioning from pre-B to immature and so the cell surface markers were those of pre-B cells, but the light chain loci had begun rearranging.

2.5.2 CDR3 Analysis

Before analysing the properties of the heavy and light chain CDR3, entries which may skew the data are removed from the full dataset, resulting in a "clean" dataset. Entries were removed from the heavy chain datasets if the CDR-H3 length was reported to be greater than 35 amino acids and entries were removed from the light chain datasets if the CDR-L3 length was reported to be greater than 20 amino acids as this was deemed biologically implausible.

Entries were removed from both the heavy and light datasets if IMGT had labelled them as unproductive because these sequences often contained frameshifts or stop codons which would interfere with the CDR3 amino acid sequence.

2.5.3 Principal Component Analysis

Principal component analysis (PCA) was calculated for all 29,447 kappa and 8,876 lambda entries in the ("clean") light chain B cell Development dataset using the "prcomp" function in R[30]. The 10 Kidera factors (Table 2-19, page 94) of each CDR-L3 amino acid sequence were used to calculate the PCA. Data were scaled.

2.5.4 Calculating Diversity

To calculate the clonal diversity in the Ribosome Display Libraries dataset, a new dataset was created from the full dataset (i.e. all members of each clonal expansion present, not just the single, modal, sequence) which contained the clone identifiers and a column containing the sample identifiers (i.e. the information by which the dataset should be split). The number of times a clone identifier appeared in the dataset indicated the size of the clone. The R package Alakazam (Gupta et al., 2015) was used to calculate and plot clonal diversity index curves using the default settings in the "rarefyDiversity" and "plotDiversityCurve" commands.

[30] The R script used to calculate the principal component analysis (PCA) was written by Grace Lu (KCL)

2.6 Modelling Kappa/Lambda CDR Structural Differences[31]

The Protein Data Bank (PDB) (Berman et al., 2000) was filtered for the antibody structures which had been solved by X-Ray diffraction to a resolution of less than 3 Å. The antibody structure had to have the heavy and the light chain. SAbDab (Dunbar et al., 2014) was used to build kappa and lambda datasets of the extracted structures. Redundancy in the PDB structures was eliminated by culling the structures using PISCES (Wang & Dunbrack, 2003) to a maximum mutual sequence identity of 99%.

199 kappa and 106 lambda structures remained and were used for analysis. Due to incomplete CDR information in six PDB structures (4LSQ, 4OB5, 4Y5Y, 4HKX, 5D70, and 7FAB), kappa antibody CDR-H1, lambda antibody CDR-H2, and lambda antibody CDR-L2 analyses were performed using 197, 105, and 103 entries respectively.

For this analysis the CDR regions were defined using the Chothia definition (Chothia & Lesk, 1987, Al-Lazikani et al., 1997). Secondary structure probabilities were calculated for each of the heavy and light chain CDR regions using DSSP (Kabsch & Sander, 1983). The hydrogen bond definition energy cut-off was set to less than -0.5 kcal/mol. The secondary structure probabilities were normalised according to CDR length. The structural definitions were as set out in Table 2-21.

Statistical significance was calculated in R using bootstrapping (boot function). 100 randomly resampled subsets were generated for each reference dataset. The mean frequency of each structure (Table 2-21) was plotted with error bars representing the 95% confidence intervals of the bootstrapped distributions. Datasets were considered significantly different if the 95% confidence intervals did not overlap.

Table 2-21: Definitions of CDR loop structural groups used in the investigation of whether kappa and lambda antibody CDR loops are structurally different.

Structure	Definition
Beta	β-strands and β-bridges
Coil	Random coil
Helix	α-helices, 3_{10}-helices and π-helices
Turn	3, 4 and 5 turns and non-hydrogen bonded bends

[31] All of the modelling and statistical analysis described in section 2.6 was performed by Julie Laffy (KCL)

2.7 Antibody Cloning and Expression: Restriction Digest Method

2.7.1 Cloning Immunoglobulin Variable Regions into PEU Vectors

Antibody heavy and light chain variable region nucleotide sequences were ordered from Integrated DNA Technologies with restriction sites at either end to enable their insertion into expression vectors. The IgH variable region sequences had a BssHII site at the 5' end and a BstEII site at the 3' end. The IgL variable region sequences had an ApaLI site at the 5' end and a PacI site at the 3' end. The DNA was ordered in a lyophilised state and was reconstituted with DNase-free H_2O to a concentration of 10 ng/µl.

The IgH and IgL inserts underwent a restriction digest as detailed in Table 2-22. The restriction digest was incubated at 50°C for 1 h, then 60°C for 1 h.

Table 2-22: Immunoglobulin variable region restriction digest.

Heavy Chain Reaction Reagent	Quantity	Light Chain Reaction Reagent	Quantity
IgH insert DNA (10 ng/µl)	40 µl (400 ng)	IgL insert DNA (10 ng/µl)	40 µl (400 ng)
BssHII[32]	1 µl	ApaLI[35]	1 µl
BstEII[33]	1 µl	PacI[36]	1 µl
NEBuffer 3.1[34]	5 µl	NEBuffer 1.1[37]	5 µl
dH_2O	3 µl	dH_2O	3 µl
Total	50 µl	Total	50 µl

The digested variable region inserts were purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) following the manufacturer's instructions. The fragments were eluted in 30 µl dH_2O.

The expression vectors were linearized using a restriction digest as described in Table 2-23. PEU1.3[38] vector digestions were incubated at 50°C for 30 min, then increase to 60°C for 30 min. PEU3.4 and PEU4.4 plasmid digestions were incubated for 37°C for 1 h.

[32] BssHII, R0199S, New England Biolabs
[33] BstEII, R0162S, New England Biolabs
[34] NEBuffer 3.1, B7203S, New England Biolabs
[35] ApaLI, R0507S, New England Biolabs
[36] PacI, R0547S, New England Biolabs
[37] NEBuffer 1.1, B7201S, New England Biolabs
[38] The vectors PEU1.3, PEU3.4 and PEU4.4 are MedImmune property

Table 2-23: IgG expression plasmid restriction digest.

Heavy Chain Reaction Reagent	Quantity	Light Chain Reaction Reagent	Quantity
		PEU3.4 kappa vector (7.0 kb)	13 µl (5 µg)
		or	
PEU1.3 vector (9.3 kb)	30 µl (15 µg)	PEU4.4 lambda vector (6.8 kb)	13 µl (5 µg)
BssHII[39] (4000 U/ml)	8 µl	ApaLI[42] (10,000 U/ml)	2 µl
BstEII[40] (10,000 U/ml)	8 µl	PacI[43] (10,000 U/ml)	2 µl
NEBuffer 3.1[41]	6 µl	NEBuffer 1.1[44]	2 µl
dH_2O	8 µl	dH2O	1 µl
Total	60 µl	Total	20 µl

The linearized vectors were run at 110 V for 30 min on a 1% (w/v) TAE agarose gel stained with SYBR SAFE (S33102, Invitrogen). The bands were excised (9.3kb, 7.0kb and 6.8kb for PEU1.3, PEU3.4 and PEU4.4 respectively) and purified using the NucleoSpin Gel and PCR Clean-up kit (740609, Macherey-Nagel) following the manufacturer's instructions. The plasmid DNA was eluted in 30 µl dH_2O and the concentration was measured using NanoDrop.

The digested variable region inserts were ligated into the relevant linearized expression vectors using the ligation reaction described in Table 2-24. The ligation reaction was incubated at 25°C for 10 min.

Heavy chain variable regions were cloned into the PEU1.3 vector upstream of an IgG1 heavy chain constant region. Kappa light chain variable regions were cloned into the PEU3.4 vector upstream of a kappa light chain constant region. Lambda light chain variable regions were cloned into the PEU4.4 vector upstream of a lambda light chain constant region.

Table 2-24: Ligating variable region inserts into IgG expression plasmids. Plasmids and inserts were in a 1:3 molar ratio.

Reagent	Quantity
PEU1.3/PEU3.4/PEU4.4 plasmid (Table 2-23)	100 ng
Variable region insert (Table 2-22)	3-fold molar excess
2X Quick Ligation Reaction Buffer[45]	15 µl
Quick T4 DNA Ligase[45]	1.5 µl
dH_2O	To 30 µl
Total	30 µl

[39] BssHII, R0199S, New England Biolabs
[40] BstEII, R0162S, New England Biolabs
[41] NEBuffer 3.1, B7203S, New England Biolabs
[42] ApaLI, R0507S, New England Biolabs
[43] PacI, R0547S, New England Biolabs
[44] NEBuffer 1.1, B7201S, New England Biolabs
[45] Quick Ligation Kit, B2200S, New England Biolabs

Plasmids were transformed into *E. coli* by adding between 1 µl and 10 µl of the ligation reaction to 100 µl DH5α competent *E. coli*. The cells were incubated on ice for 5 min and then spread on 2X TY agar containing 100 µg/ml ampicillin, 2% (w/v) glucose. The plates were incubated overnight at 37°C.

Colony PCR was used to check for the presence of inserts. 10 µl of PCR master mix (see Table 2-25) was added to the required number of wells on a PCR plate. Individual colonies were picked and were transferred both to a PCR tube and 150 µl 2X TY broth containing 100 µg/ml ampicillin, 2% (w/v) glucose. The cultures were incubated at 37°C, 150 rpm until the results of the colony PCR indicated which contained variable region inserts. The colony PCR was run using the conditions detailed in Table 2-25.

Table 2-25: IgG expression plasmid colony PCR reagents and thermal cycle. Quantities for 1 ml of master mix are given. Primer 3.4PCRrev was used for PEU3.4 vectors. Primer 4.4PCRrev was used for PEU4.4 vectors. All primers in this table are MedImmune property.

Heavy Chain Reaction		Light Chain Reaction	
Reagent	Quantity	Reagent	Quantity
2X ReddyMix PCR Master Mix[46]	500 µl	2X ReddyMix PCR Master Mix[46]	500 µl
pEUPCRfor primer (100 µM)	1 µl	pEUPCRfor primer (100 µM)	1 µl
		3.4PCRrev primer (100 µM)	1 µl
		or	
IgGHCrev primer (100 µM)	1 µl	4.4PCRrev primer (100 µM)	1 µl
dH_2O	498 µl	dH_2O	498 µl
Total	1 ml	Total	1 ml

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	95°C	2 min	X1
Denaturation	94°C	30 s	X30
Annealing	50°C	30 s	
Extension	72°C	90 s	
Final extension	72°C	5 min	X1
Hold	4°C	∞	

The 10 µl PCR product was run on a 1% (w/v) agarose gel at 110 V for 10 min. The size of the bands indicated which ligations had worked. One successful ligation per insert was cultured in 5 ml 2X TY broth containing 100 µg/ml ampicillin, 2% (w/v) glucose overnight at 37°C, 150 rpm. The colonies were sent for sequencing to confirm the presence of the correct insert.

[46] 2x ReddyMix PCR Master Mix, AB-0575/DC/LD, Thermo Scientific

2.7.2 Expression of IgG1 Antibodies from PEU Vectors[47]

Plasmids containing the desired inserts (section 2.7.1) were transfected into CHO cells for expression as IgG1 antibodies. The antibodies were purified in PBS using a MabSelect SuRe affinity chromatography column (GE Healthcare) on an ÄKTA Pure protein purification system. SDS-PAGE was used to check the successful expression of complete heavy and light chains from each transfection. HP-SEC was used to measure the proportion of IgG1 monomers in the purified sample and any aggregation or fragmentation.

[47] The work described in section 2.7.2 was performed by the IgG team at MedImmune (Cambridge, United Kingdom)

2.8 Antibody Cloning and Expression: PIPE Cloning Method

Antibody variable region sequences were cloned into modified pVITRO1-hygro-mcs (Invivogen) expression vectors (see Figure A-1, page 270, for plasmid maps) using polymerase incomplete primer extension (PIPE) as described in Dodev et al. (2014).

The principle of PIPE cloning is to amplify the two halves of the expression plasmid and the inserts using primers with complementary overhangs. The PCR used has a short extension time (55 s for plasmid halves, 5 s for inserts), ensuring that the extension is not always complete. This means that there is a single-stranded 5' overhang at either end of some PCR products where the polymerase has not reached the end. As the primers for amplifying the vector halves and inserts were complementary, the 5' overhangs are complementary. This means that the plasmid halves and inserts can simply be mixed, the complementary ends will ligate, and complete plasmids will assemble (see Figure 2-5 for schematic). Unlike the plasmids used in section 2.7, the heavy and light chain pair are inserted into the same pVITRO1 expression vector.

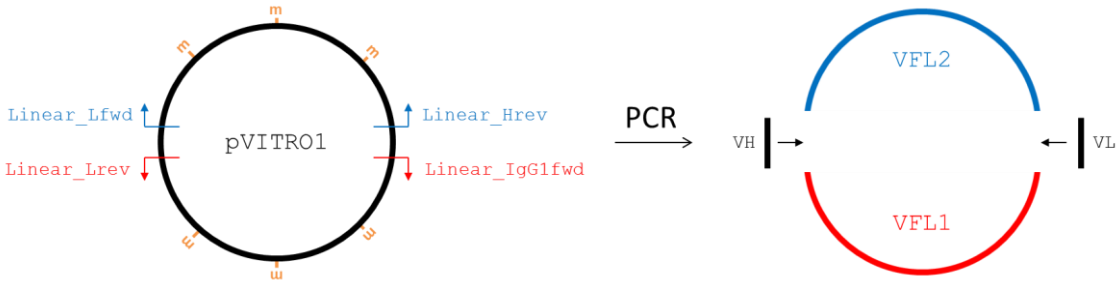


Figure 2-5: PIPE cloning principle. Whole, circular, pVITRO1 plasmid is isolated from *E. coli* stocks. The plasmid is copied in two halves and the VH and VL inserts are amplified using primers with complementary overhangs. The short extension time ensures some incomplete PCR amplification, resulting in plasmid halves and inserts with complementary 5' single-stranded overhangs. The plasmid halves and inserts are mixed and they ligate to form complete plasmids. The original plasmid not containing insert is removed using a DpnI digest as only original plasmid will be Dam methylated. The lambda pVITRO1 plasmid has been used in this example; see Figure A-1 (page 270) for pVITRO1 plasmid maps.

2.8.1 pVITRO1 Vector Linearization by PCR

The pVITRO1-dV-IgG1_κ and pVITRO1-dV-IgG1_λ plasmids were copied into two linear halves (VFK1 + VFK2 and VFL1 + VFL2 respectively) by PCR. See Table 2-26 for vector linearization primers and Table 2-27 for PCR reagents and conditions.

Table 2-26: pVITRO1 vector linearization primers. See Table A-15 (page 268) for primer sequences.

Vector template	Forward primer	Reverse primer	Vector fragment
pVITRO1-dV-IgG1_κ	Linear_IgG1fwd	Linear_Krev	VFK1
	Linear_Kfwd	Linear_Hrev	VFK2
pVITRO1-dV-IgG1_λ	Linear_IgG1fwd	Linear_Lrev	VFL1
	Linear_Lfwd	Linear_Hrev	VFL2

Table 2-27: pVITRO1 vector linearization PCR reagents and thermal cycle. The vector templates should be purified from *E. coli* culture.

Reagent	Quantity
Vector template pVITRO1-dV-IgG1_κ/λ (10 ng/µl)	1 µl
Forward primer (10 µM) (see Table 2-26)	2.5 µl
Reverse primer (10 µM) (see Table 2-26)	2.5 µl
H_2O	19 µl
Phusion Flash High-Fidelity PCR Master Mix (2X)[48]	25 µl
Total	50 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X30
Annealing	60°C	15 s	
Extension	72°C	55 s	NO FINAL
Hold	4°C	∞	EXTENSION

1 µl of amplified vector product was run on a 0.8% TAE agarose gel to confirm linearization. The bands should be approximately 4 kb in size.

To minimise background colonies, a DpnI digest (Table 2-28) was performed on the linearized vector fragments to eliminate whole circular template plasmid. The DpnI digest was incubated at 37°C for 1 h.

Table 2-28: pVITRO1 linearized vector DpnI digest.

Reagent	Quantity
Linearized vector PCR product (Table 2-27)	49 µl
CutSmart Buffer (10X)[49]	6 µl
DpnI[50]	1 µl
H_2O	4 µl
Total	60 µl

[48] Phusion Flash High-Fidelity PCR Master Mix (F548S, Thermo Scientific)
[49] CutSmart Buffer (B7204S, New England Biolabs)
[50] DpnI (R0176S, New England Biolabs)

2.8.2 PCR Amplification of Variable Region Sequences

Antibody variable region sequences were amplified from starting material using the PCR described in Table 2-29. Forward and reverse primers were specific to the first and last five amino acids of the variable and joining regions. See Table A-17 (page 269) for the primer combinations used to amplify the heavy and light chain variable regions of each antibody expressed using this method.

Table 2-29: Insert amplification PCR reagents and thermal cycle. Template DNA is the variable region sequence. See Table A-16 (page 268) for primer sequences.

Reagent	Quantity
Template DNA	1 µl
Forward primer (10 µg/ml)	1 µl
Reverse primer (10 µg/ml)	1 µl
H_2O	7 µl
Phusion Flash High-Fidelity PCR Master Mix (2X)[51]	10 µl
Total	20 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X30
Annealing	60°C	15 s	
Extension	72°C	5 s	NO FINAL
Hold	4°C	∞	EXTENSION

1 µl of each PCR product was run on a 1.3% TAE agarose gel to check for bands of the correct size (IgH ~400 bp, IgK/IgL ~350 bp).

2.8.3 Vector Assembly and Transformation

The DpnI-treated vector fragments (section 2.8.1) were mixed with the inserts (section 2.8.2) in the order and quantities shown in Table 2-30 and incubated at room temperature for 1 h. No heating or cooling was required.

Table 2-30: pVITRO1 vector assembly reaction with inserts.

Order	Kappa	Lambda	Quantity
1	H_2O	H_2O	4 µl
2	VFK1 plasmid half	VFL1 plasmid half	1 µl
3	VK variable region	VL variable region	1 µl
4	VFK2 plasmid half	VFL2 plasmid half	1 µl
5	VH variable region	VH variable region	1 µl

[51] Phusion Flash High-Fidelity PCR Master Mix (F548S, Thermo Scientific)

2 µl of vector assembly reaction mix (Table 2-30) was transformed into 25 µl NEB 10-beta Competent *E. coli* (C3019H, New England Biolabs). The transformation protocol was as per the manufacturer's instructions with the single modification that the *E. coli* were re-suspended in 475 µl SOC media and then incubated for 90 min at 37°C, 200 rpm. 200 µl of transformation reaction was spread on Fast-Media Hygro Agar (InvivoGen) or Fast-Media Base Agar (InvivoGen) supplemented with 100 µg/ml Hygromycin B Gold (InvivoGen) and incubated overnight at 37°C.

Colonies were screened for the presence of a heavy chain variable region insert using colony PCR (Table 2-31). Each colony was saved on a patch plate (selective hygromycin-containing agar as described in previous paragraph). In every case, if a heavy chain variable region was present in the plasmid, the light chain variable region insert was also present in the plasmid.

15 µl of colony PCR products were run on a 1.3% (w/v) TAE agarose gel to confirm the presence of insert. A band of approximately 1.7 kb indicates the presence of a heavy chain variable region. A band of approximately 1.4 kb indicates the absence of a heavy chain variable region.

Table 2-31: pVITRO1 colony PCR reagents and thermal cycle. See Table A-18 (page 269) for primer sequences.

Reagent	Quantity
Single colony	-
pVITRO1F primer (10 µg/ml)	1 µl
pVITRO1R primer (10 µg/ml)	1 µl
H_2O	5.5 µl
One*Taq* Quick-Load 2X Master Mix[52]	7.5 µl
Total	15 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	94°C	3 min	X1
Denaturation	94°C	15 s	X30
Annealing	60°C	15 s	
Extension	68°C	2 min	
Hold	4°C	∞	

Colonies containing successfully assembled plasmids were cultured in 5 ml Fast-Media Hygro TB (InvivoGen) or Fast-Media Base LB (InvivoGen) supplemented with 100 µg/ml Hygromycin B Gold (InvivoGen). Plasmids were purified using the QIAprep Spin Miniprep Kit (27104, Qiagen) following the manufacturer's instructions. Plasmids were eluted in 50 µl H_2O and samples were sent for Sanger sequencing (Eurofins, Germany) to confirm the successful insertion of the antibody variable regions (see Table A-18, page 269, for sequencing primers).

[52] One*Taq* Quick-Load 2X Master Mix with Standard Buffer (M0486S, New England Biolabs)

2.8.4 Expression of IgG1 Antibodies from pVITRO1 Vectors

Antibodies were either expressed in adherent HEK 293 cells or in FreeStyle 293-F cells in suspension.

When antibodies were expressed in HEK 293 cells, cells were seeded in 12-well tissue culture plates at 1×10^5 cells/ml in 2 ml complete media (DMEM[53], 10% (v/v) HI-FCS, 1% (v/v) penicillin-streptomycin). Cells were grown for 24 h before transfection. Cells were transfected using FuGENE HD transfection reagent (E2311, Promega) following manufacturer's instructions in a 1:3 DNA:FuGENE ratio. 2 µg DNA was used per well. Cells were incubated at 37°C, 5% CO_2 for up to one week. Tissue culture supernatant was clarified by centrifugation in a benchtop centrifuge at maximum speed for 5 min.

When antibodies were expressed in FreeStyle 293-F cells (Gibco), cells were seeded in suspension in a six-well tissue culture plate at a concentration of 1×10^6 cells/ml in 2 ml FreeStyle 293 Expression Medium (12338018, Gibco). Cells were immediately transfected with 2 µg DNA as follows: DNA and PEI Max were separately diluted in Opti-MEM Reduced Serum Media as described in Table 2-32. The two solutions were then mixed together and incubated at room temperature for 20 minutes. The plasmid/PEI Max/Opti-MEM mixture (200 µl) was then gently added to the 2 ml of FreeStyle 293-F cells. The transfected cells were incubated at 37°C, 7% CO_2, 135 rpm for five days. Tissue culture supernatant was clarified by centrifugation at 1500 g for 10 min.

In both methods, the concentration of IgG1 in the clarified tissue culture supernatant was measured using a quantitative ELISA (section 2.8.5).

Table 2-32: DNA and PEI Max dilutions in OptiMEM. pVITRO1 DNA (µg) and PEI Max (µl) were used in a 1:3 ratio. The total of both solutions when mixed together was 200 µl, i.e. 10% of the FreeStyle 293-F cell culture volume.

DNA solution		PEI Max solution	
pVITRO1 plasmid DNA	2 µg	PEI Max (1 mg/ml)[55]	6 µl
Opti-MEM[54]	To 100 µl	Opti-MEM[54]	94 µl

[53] DMEM, high glucose, GlutaMAX, pyruvate (31966, Thermo Fisher Scientific)
[54] Opti-MEM Reduced Serum Media, 31985062, Gibco
[55] Polyethylenimine Hydrochloride, Linear (MW 4,000) (24885, Polysciences Inc.) at 1 mg/ml in H_2O

2.8.5 Quantitative ELISA

A Nunc MaxiSorp 96-well ELISA plate was coated with 50 µl goat anti-human IgG-UNLB (204001, Oxford Biotech) at 2 µg/ml in PBS. The plate was incubated at 4°C overnight. The plate was washed three times with 0.05% (v/v) Tween-20 in PBS and blocked with 200 µl 3% (w/v) skimmed milk powder in PBS for 1 h at room temperature. The plate was washed three times as before. To create the standard curve, rows A-C, columns 1-11, were coated with 50 µl of two-fold serial dilutions of IgG from Human Serum (I4506, Sigma Aldrich) in PBS. The initial concentration in column one was 3 µg/ml. The concentration in column 11 was 0.003 µg/ml. Column 12 contained PBS only. The rest of the plate was coated with 50 µl of clarified tissue culture supernatant (section 2.8.4) in various dilutions in PBS. The plate was incubated for 1 h at room temperature and then washed three times as before. The plate was then coated with 50 µl Anti-Human IgG (Fc specific)-Peroxidase antibody produced in goat (A0170, Sigma Aldrich) in a 1/10,000 dilution in blocking buffer. The plate was incubated for 1 h at room temperature. The plate was washed three times as before. The plate was then coated with 50 µl OPD substrate[56] and incubated for 30 min at room temperature in the dark. The reaction was stopped by the addition of 50 µl 3 M HCl. The plate was read at 492 nm.

The concentration of antibodies in clarified supernatant was determined by finding the dilution at which the optical density of the antibody of interest intersects with the linear region of the standard curve. The concentration of the antibody can then be read off the graph as shown in Figure 2-6.

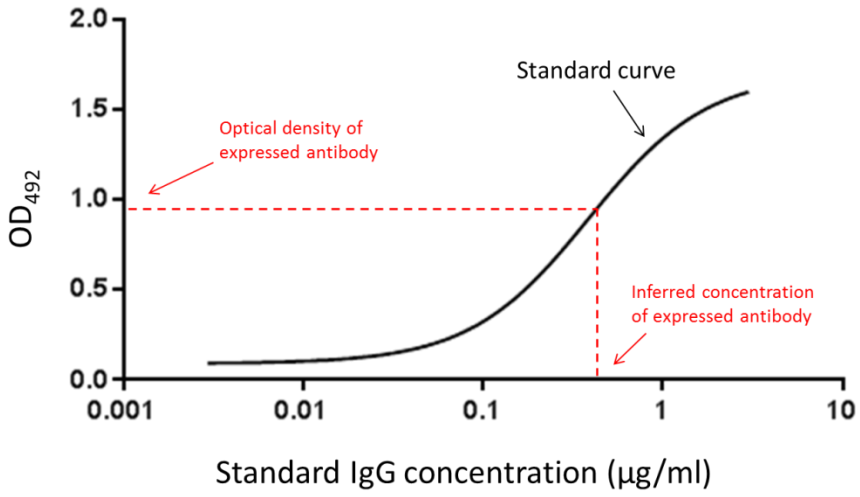


Figure 2-6: Estimating IgG concentration from quantitative ELISA standard curve. Standard IgG is titrated in a series of 1-in-2 dilutions across the top 11 columns of a 96-well ELISA plate to create the standard curve. The concentration of expressed antibodies is then estimated by reading off the concentration of the standard IgG at the relevant optical density.

[56] One 5 mg OPD tablet (Sigma, P6912-50 TAB) dissolved in 10 ml 1X peroxide substrate buffer (9 ml H_2O, 1 ml 10X stable peroxide substrate buffer (Thermo Scientific, 34062))

2.9 HEp-2 Immunocytochemistry

2.9.1 HEp-2 Cell Culture on Glass Coverslips

HEp-2 cells (CCL-23, ATCC) were seeded at 2 X 10^5 cells/ml in 500 µl of DMEM (41966-029, Gibco), 10% (v/v) HI-FCS, 1% (v/v) penicillin-streptomycin in 24-well tissue culture plates, each well contained a single sterile glass 13 mm coverslip (631-0150, VWR). The cells were incubated at 37°C, 5% CO_2 until the cells reached 70% confluency on the coverslips.

2.9.2 Immunofluorescence Staining of HEp-2 Cells

The tissue culture supernatant was removed from each well and the coverslips were washed three times in PBS. The coverslips were incubated with 500 µl 4% (v/v) paraformaldehyde (PFA) in PBS at room temperature for 10 min. The coverslips were washed three times in PBS and incubated in 500 µl of 0.1% (v/v) Triton X-100, 3% (w/v) BSA in PBS at room temperature for 10 min. The coverslips were washed once in PBS and incubated in 300 µl 50 µg/ml primary antibody diluted in 3% (w/v) BSA in PBS at room temperature for 1 h. The coverslips were washed three times in PBS and incubated for 1 h at room temperature in 300 µl 1/500 goat anti-human IgG (H+L) Alexa Fluor 488 conjugate (polyclonal, A-11013, Thermo Fisher) and 1/10,000 DAPI (D9542-10MG, Sigma Aldrich) diluted in 3% (w/v) BSA in PBS. The coverslips were washed three times in PBS, carefully removed from the 24-well plate and mounted on glass microscope slides using 5 µl Glycergel mounting medium (C0563, Dako). Slides were stored in the dark at 4°C until use.

2.9.3 Fluorescent Imaging

Slides were imaged using 3 µm Z-stacks on a Nikon Eclipse Ti widefield microscope. The exposure time for both the DAPI and Alexa Fluor 488 was 100 ms. The presence of DAPI-stained nuclei was used to confirm that there were cells present in every picture. The ND2 image files were analysed using ImageJ (Rasband, 1997-2016).

2.10 Solid-Phase ELISA

96-well Nunc Maxisorp ELISA plates (M9410, Sigma Aldrich) were coated with 50 µl of antigen diluted in PBS to the optimal coating concentration (Table 2-34)[57]. The plates were incubated overnight at 4°C. Plates were washed three times with washing buffer (see Table 2-33) and then incubated with 200 µl-300 µl blocking buffer (Table 2-33) for 1 h at room temperature. The plate was then washed three times in washing buffer (Table 2-33). 50 µl of primary test antibody (IgG1) was added at a concentration of 1 µg/ml in PBS. 50 µl of primary positive control antibody in PBS (see Table 2-35 for working dilutions) was included on every plate. The primary antibodies were incubated under the conditions described in Table 2-33. The plates were washed three times with washing buffer (Table 2-33) and 50 µl of secondary, HRP-conjugated, antibody diluted in blocking buffer (see Table 2-35 for working dilutions and Table 2-33 for blocking buffers). The plates were then incubated as described in Table 2-33. The plates were washed three times with washing buffer (Table 2-33) and 50 µl OPD substrate[58] or TMB substrate (T0440, Sigma Aldrich) was added per well. The plates were incubated on the bench until the positive controls changed colour. The reactions were stopped with 50 µl 3 M HCl (OPD substrate) or 0.5 M H_2SO_4 (TMB substrate). The plates were read at 492 nm (OPD substrate) or 450 nm (TMB substrate).

Table 2-33: Blocking, washing and incubation conditions used in solid-phase ELISA. Blocking and washing condition combinations are shown together.

Blocking conditions		
Blocking buffer		Washing buffer
3% (w/v) Marvell skimmed milk powder in 1X PBS		1X PBS
3% (w/v) Bovine serum albumin (BSA) in 1X PBS		1X PBS
0.05% (v/v) Tween-20, 2 mM EDTA in 1X PBS		0.05% (v/v) Tween-20, 2 mM EDTA in 1X PBS
50 mM Tris, 1% (v/v) Tween-20, 0.3 M NaCl, pH 7.5		50 mM Tris, 1% (v/v) Tween-20, 0.3 M NaCl, pH 7.5
Incubation conditions		
Temperature	Agitation	Time (primary/secondary Ab)
37°C	200 rpm	2 h/1.5 h
Room temperature	No agitation	1 h/1 h

[57] The optimal coating concentration for each antigen was determined by titration curves (Figure A-5, page 284)

[58] One 5 mg OPD tablet (Sigma, P6912-50 TAB) dissolved in 10 ml 1X peroxide substrate buffer (9 ml H_2O, 1 ml 10X stable peroxide substrate buffer (Thermo Scientific, 34062))

Table 2-34: Optimal coating concentrations for antigens used in solid-phase ELISA. Antigen coating concentrations were determined by titration curves (see Figure A-5, page 284).

Antigen	Accession #	Catalogue #	Manufacturer	Coating concentration
dsDNA		15633-019	Invitrogen	2.5 µg/ml
Insulin		I5523	Sigma	2.5 µg/ml
GIF	P27352	8387-TC	R&D Systems	5 µg/ml
LPS		L7770	Sigma	10 µg/ml
MPO	P05164	3174-MP	R&D Systems	10 µg/ml
ApoH	P02749	5087-AH	R&D Systems	2.5 µg/ml

Table 2-35: Working dilutions of commercial antibodies used in solid-phase ELISA. Positive control primary antibodies and HRP-conjugated secondary antibodies are shown.

Primary antibody	Clone	Catalogue #	Manufacturer	Working dilution
Anti-ds DNA	35I9 DNA	ab27156	AbCam	1/2000
Anti-Insulin	K36AC10	I2018	Sigma Aldrich	1/2000
Anti-Salmonella LPS	D46J	MA5-18257	Thermo Fisher	1/200
Anti-ApoH	517038	MAB5087	R&D Systems	1/500
Anti-GIF	139.1B8	MCA5885G	AbD Serotec	1/1000
Anti-MPO	Polyclonal	ab45977	AbCam	1/1000
Secondary antibody	**Clone**	**Catalogue #**	**Manufacturer**	**Working dilution**
Anti-Human IgG (Fc specific)-Peroxidase	Polyclonal	A0170	Sigma Aldrich	1/10,000
Anti-Mouse IgG (Fc specific)-Peroxidase	Polyclonal	A2554	Sigma Aldrich	1/5000
Anti-Rabbit Ig-Peroxidase	RG-16	A2074	Sigma Aldrich	1/5000

2.11 Soluble ELISA

Dynabeads M-280 Streptavidin (11205D, Thermo Fisher) were re-suspended and a total volume of 10 µl per reaction was transferred into two clean tubes (one for the experiment, one as a control). The beads were pelleted using a magnetic separator, the supernatant was discarded and the beads were re-suspended in the same volume of PBS. This wash step was repeated twice more. The beads were then re-suspended in the same volume of blocking buffer (3% (w/v) BSA in PBS) and incubated on a rotary mixer (20 rpm) for 1 h at room temperature. The beads were washed three times in PBS as described previously. The beads in one tube were re-suspended in 10X the original liquid volume of 0.1 µg/ml biotinylated insulin[59] in blocking buffer. The beads in the control tube were re-suspended in the same volume of blocking buffer only. The tubes were incubated on a rotary mixer (20 rpm) for 1 h at room temperature. The beads were washed three times in PBS and, on the third wash, 10 µl was transferred into the required number of wells in a 96-well PCR plate. The plate was placed on a magnetic separator to pellet the beads and the 10 µl of PBS was carefully removed using a pipette. Without letting the beads dry out, the beads were re-suspended in 50 µl of 1 µg/ml primary antibody in PBS. As a positive control, 50 µl of anti-insulin antibody was used at the concentration stated in Table 2-35 (page 111). As a negative control, 50 µl PBS was used. The plate was incubated at 37°C, 200 rpm for 2 h. The beads were washed three times in 50 µl PBS as described previously. The beads were re-suspended in 50 µl secondary antibody in blocking buffer at the concentrations stated in Table 2-35. The plate was incubated at 37°C, 200 rpm for 1.5 h. The beads were washed three times in PBS and, on the last wash, the beads were transferred to a clear, round-bottomed 96-well plate. The plate was placed on a magnetic separator and the PBS was removed. The beads were re-suspended in 50 µl TMB substrate (T0440, Sigma Aldrich) and the plate was incubated at room temperature until the positive controls had changed colour. The reaction was stopped by adding 50 µl 0.5 M H_2SO_4. The plate was read at 450 nm.

This method is illustrated in a schematic in Figure 3-15 (page 168).

[59] The biotinylated insulin was made in-house at MedImmune (Cambridge, United Kingdom); biotin was attached to the lysines on the insulin. A titration curve was used to determine that 0.1 µg/ml was sufficient to saturate the streptavidin-coated Dynabeads (see Figure A-6, page 284).

2.12 Thermal Shift Assay

A thermal shift assay measures the temperature at which a protein denatures, exposing hydrophobic patches which are usually hidden when the protein is in its native state.

20 µl of each purified antibody was added in triplicate to a white PCR plate. 5 µl 40X SYPRO Orange (S6650, Thermo Fisher Scientific) was added to each well (8X final concentration). An RT-PCR machine was used to perform a melting curve. The temperature started at 20°C and was increased by 1°C (holding at each temperature for 1 s) up to a final temperature of 90°C. The fluorescence was read at every 1°C increment. The melting temperature (Tm) could be determined from the point at which a fluorescent signal was detected.

This method is illustrated in a schematic in Figure 3-22A (page 174).

2.13 Phage Display: Panning Selection against Multiple Antigens

2.13.1 Panning Selections

One well of a Nunc Maxisorp plate was coated with 100 µl antigen at an appropriate concentration (Table 2-34, page 111) and incubated overnight at 4°C. The plate was washed three times with PBS and incubated with 300 µl blocking buffer (3% (w/v) skimmed milk powder in PBS) for 1 h at room temperature and then washed once with PBS.

A single 50 µl aliquot of BMVtrp phage library (10^{12} M13 phage)[60] was blocked by adding 50 µl 2X blocking buffer (6% (w/v) skimmed milk powder in 2X PBS) and incubating for 1 h at room temperature.

The blocked phage library was first transferred to a blank ELISA well (coated with PBS only) which had been blocked with 3% (w/v) skimmed milk powder in PBS and incubated for 1 h at room temperature. The contents (i.e. phage that had not bound) was then transferred to the antigen-coated well and incubated for 1 h at room temperature. The well was washed five times with PBS to remove the unbound phage. To remove the antigen-bound phage, 100 µl trypsin solution (10 µg/ml trypsin in 0.1 M NaP buffer, pH 7.0) was added to the well and incubated at 37°C, 200 rpm for 30 min.

The eluted phage were transferred to 0.8 ml mid-log phase TG1 *E. coli* and incubated at 37°C, 200 rpm for 1 h. This allowed the eluted phage to infect the bacteria.

2.13.2 Isolating Selected Phage and Calculating Output Titre

To gauge the success of the selection, output titres were calculated. Output titres were created by serially diluting the M13 phage-infected TG1 *E. coli* from section 2.13.1, plating out the diluted *E. coli* and counting the resulting colonies.

10 µl of the infected TG1 *E. coli* culture was diluted in 90 µl 2X TY broth and mixed thoroughly (10^{-1} dilution). 10 µl of infected TG1 *E. coli* was diluted in 990 µl 2X TY and mixed thoroughly (10^{-2} dilution). 100 µl of the 10^{-2} dilution was diluted in 900 µl 2X TY broth and mixed thoroughly (10^{-3} dilution). 100 µl of the 10^{-3} dilution was diluted in 900 µl 2X TY broth and mixed thoroughly (10^{-4} dilution). 100 µl of each dilution (10^{-1}, 10^{-2}, 10^{-3} and 10^{-4}) was plated on a 2X TY agar plate supplemented with 100 µg/ml ampicillin and 2% (w/v) glucose. The cultures were grown at 30°C overnight.

[60] The BMVtrp phage library was made in-house at MedImmune (Cambridge, United Kingdom). The original construction of this library was published in Vaughan et al. (1996).

The remaining *E. coli* was spread on a bioassay plate of the same media and grown at 30°C overnight.

The colonies on the output titre plates were counted and the following equation was used to determine the number of colony-forming units per millilitre (cfu/ml):

$$\frac{No.of\ colonies \times Dilution\ factor}{Volume\ plated\ (ml)} = No.of\ colony\ forming\ units\ per\ ml\ (cfu/ml)$$

After calculating the output titre, individual output titre colonies could be picked for sequencing and testing the specificity of the scFv using ELISA (section 2.13.4).

If the output titre was within a reasonable range, the bacterial lawn on the bioassay plate was scraped into 5 ml 2X TY broth supplemented with 50% (v/v) glycerol in a 2:1 ratio. The bacteria were re-suspended for 10 min on an end-over-end rotator. The phage from the bioassay plate *E. coli* were used in the next round of panning selection (section 2.13.3).

2.13.3 Calculating Input Titre and Isolating Phage for Further Panning Selections

The bioassay plate *E. coli* were grown in 25 ml 2X TY broth supplemented with 100 µg/ml ampicillin and 2% (w/v) glucose. The culture was grown at 37°C, 280 rpm, to an OD_{600} of 0.5-1.0. The culture was then superinfected with 2.5 µl stock M13KO7trp helper phage (3 X 10^{13}pfu/ml) and incubated at 37°C, 150 rpm for 1 h.

The *E. coli* were pelleted at 2000 g for 10 min, re-suspended in 500 µl 2X TY broth supplemented with 100 µg/ml ampicillin and 50 µg/ml kanamycin. The re-suspended *E. coli* were then transferred to 25 ml of the same broth and were grown overnight at 25°C, 280 rpm. 1 ml of *E. coli* culture was then pelleted in a benchtop centrifuge at maximum speed for 5 min. the input phage-containing clarified supernatant was transferred to a clean 1.5 ml polypropylene tube.

Input titres were calculated by diluting 10 µl input phage in 990 µl 2X TY broth (10^{-2} dilution). The mixture was mixed thoroughly and 10 µl was diluted in 990 µl 2X TY broth (10^{-4} dilution). The mixture was mixed thoroughly and 10 µl was diluted in 990 µl exponentially growing TG1 *E. coli* (10^{-6} dilution). The culture was incubated for 1 h at 37°C, 150 rpm. 100 µl of culture was then added to 900 µl 2X TY broth and mixed by pipetting (10^{-7} dilution). 100 µl was plated on 2X TY agar supplemented with 100 µg/ml ampicillin and 2% (w/v) glucose. 100 µl of the 10^{-7} dilution culture was diluted in 900 µl 2X TY (10^{-8} dilution) and 100 µl was plated out. This was repeated to create 10^{-9} and 10^{-10} dilutions. The 10^{-7}, 10^{-8}, 10^{-9} and 10^{-10} plates were incubated overnight at 30°C.

The colonies were counted and the input titre (cfu/ml) was determined using the same equation as the output titre. If the input titre was within a reasonable range (approximately 10^{12} cfu/ml), the input phage could be used in a second round of selection (section 2.13.1).

This process can be repeated several times (typically three to four times) to select for scFv which bind the selection antigens.

2.13.4 Using Helper Phage to Release M13 Phage from Individual *E. coli* Colonies

Individual colonies from the output titre plates (section 2.13.2) were picked into 500 µl 2X TY broth supplemented with 100 µg/ml ampicillin and 2% (w/v) glucose in a 96-well format. The cultures were grown for 5 h at 37°C, 280 rpm.

Stock M13K07trp helper phage (3 X 10^{13} pfu/ml) were diluted by adding 5 µl stock helper phage per 10 ml 2X TY broth supplemented with 100 µg/ml ampicillin and 2% (w/v) glucose. 100 µl diluted helper phage was then added to each well in the now visibly turbid 500 µl 96-well liquid cultures. The cultures were incubated again for 1 h at 37°C, 150 rpm.

The plates were centrifuged in a bench-top centrifuge at 3200 rpm for 10 min at room temperature. The supernatants were discarded and 500 µl 2X TY broth supplemented with 100 µg/ml ampicillin and 50 µg/ml kanamycin was added to every well. The 96-well culture plates were incubated overnight at 25°C, 280 rpm to promote the expression of phage.

The 500 µl cultures were blocked by adding 500 µl 2X blocking buffer (6% (w/v) milk powder in 2X PBS) and incubating for 1 h at room temperature. The plates were centrifuged on a bench-top centrifuge at 3200 rpm for 5 min at room temperature. The supernatants were then used in M13 phage ELISA (section 2.13.5).

2.13.5 M13 Phage ELISA

ELISA wells were coated with 50 µl of antigen at optimal concentration in PBS (see Table 2-34, page 111) and incubated overnight at 4°C. The plates were washed 3X with PBS and blocked with 300 µl 3% (w/v) skimmed milk powder in PBS. The plates were incubated for 1 h at room temperature. The blocked ELISA wells were washed 3X with PBS and 50 µl M13 phage-containing supernatant (section 2.13.4) was added per well. The plate was incubated for 1 h at room temperature. The wells were then washed 3X with 0.1% (v/v) Tween-20 in PBS. 50 µl HRP/Anti-M13 Monoclonal Conjugate (27942101, GE Healthcare) diluted 1/5000 in 3% (w/v) skimmed milk powder in PBS was added to each well and incubated for 1 h at room temperature.

The wells were washed 3X with 0.1% (v/v) Tween-20 in PBS. 50 µl TMB substrate (T0440, Sigma Aldrich) was added per well and incubated until the positive controls changed colour. The reaction was stopped by adding 50 µl 0.5 M H_2SO_4 to each well. The plate was read on a fluorescent plate reader at a wavelength of 450 nm.

2.14 Ribosome Display: Building the DNA Libraries

Ribosome display libraries were built for each donor in the Caucasian Aid Worker Ebola survivor dataset (section 2.2.1.1, page 83). Two libraries were built per donor: one using kappa light chains, the other using lambda light chains. A schematic of how the ribosome display libraries were built is shown in Figure 2-7. The linker sequences are shown in Table 2-36.

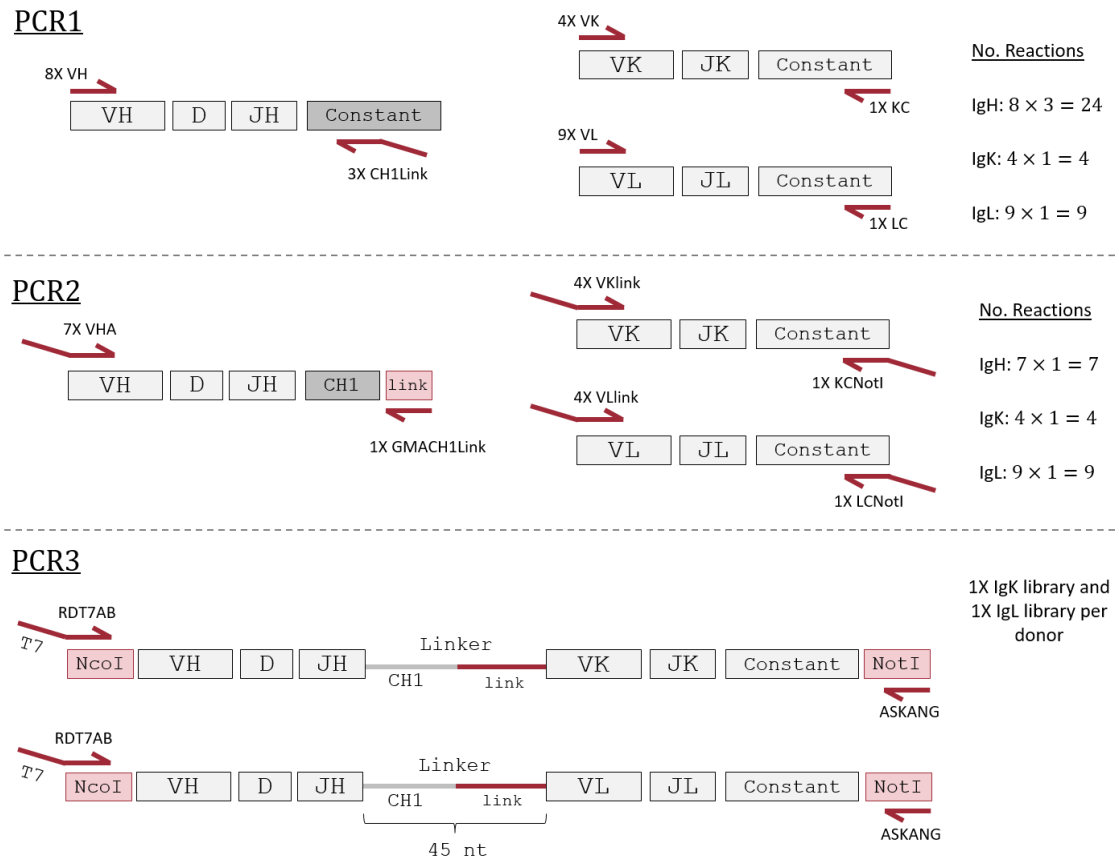


Figure 2-7: Schematic of building the ribosome display libraries. In the first PCR (PCR1), V-specific and constant region specific primers were used to amplify variable regions from cDNA. IgA, IgG, IgM, IgK and IgL sequences were amplified. Linker sequences and restriction sites were added in a second round of PCR (PCR2). Heavy and light chains were linked at random in an overlap extension PCR (PCR3) and a T7 promoter site was added to the start of each amplicon. Each linker was composed of the first 24 nucleotides from the IgA, IgG or IgM CH1 region and the remaining 21 nucleotides is the complementary overhang region from the link primers in PCR2. When translated the linker is 15 amino acids long (see Table 2-36). See section A.1.4 (page 265) for all primer sequences.

Table 2-36: Ribosome display linker sequences. In blue is the 5' region of the CH1 region which can be used to identify the heavy chain class of the original antibody. In red is the overlap sequence used to join heavy and light variable regions.

Isotype	Linker nucleotide sequence (5' to 3')	Linker amino acid sequence
IgG	gcctccaccaagggcccatcggtcttccccctggcgccctcctcc	ASTKGPSVFPLAPSS
IgM	gcatccccgaccagccccaaggtcttccccctggcgccctcctcc	ASPTSPKVFPLAPSS
IgA	gggagtgcatccgccccaaccctttccccctggcgccctcctcc	GSASAPTLFPLAPSS

2.14.1 Reverse Transcription of mRNA from Total RNA

Two reverse transcription reactions were set up per donor as shown in Table 2-37. One reaction used oligo-dT primers (thought to be optimal for light chain variable region sequences as they are shorter) and the other reaction used random hexamers (thought to be optimal for heavy chain variable region sequences as they are longer).

Table 2-37: Ribosome display reverse transcription reagents for a single donor.

Tube 1 (optimal for light chains)		Tube 2 (optimal for heavy chains)	
Reagent	Quantity	Reagent	Quantity
Oligo-d(T)$_{23}$ VN (50 µM)[61]	6 µl	Random Primer Mix (60 µM)[61]	6 µl
Total RNA	1 µg	Total RNA	1 µg
H_2O	to 8 µl	H_2O	to 8 µl
Total	8 µl	Total	8 µl

Both reactions (Table 2-37) were heated at 70°C for 5 min (denaturing) and then incubated on ice for 30 s. The contents of both tubes were pooled into a single PCR tube (16 µl) and a reverse transcription was set up as detailed in Table 2-38.

Table 2-38: Ribosome display reverse transcription reagents and incubation.

Reagent	Quantity
Pooled RNA from Table 2-37	16 µl
ProtoScript II Reaction Mix (2X)[61]	20 µl
ProtoScript II Enzyme Mix (10X)[61]	4 µl
Total	40 µl

Thermal Cycle Step	Temperature	Time	Cycles
Annealing	25°C	7 min, 30 s	X1
Extension	42°C	60 min	
Enzyme Inactivation	95°C	5 min	
Hold	4°C	∞	

[61] ProtoScript II First Strand cDNA Synthesis Kit (E6560S, New England Biolabs)

2.14.2 Amplification of Variable Regions (PCR1)

The cDNA from section 2.14.1 was used in a PCR reaction (PCR1) to amplify immunoglobulin sequences. A selection of V family-specific forward primers and constant region-specific reverse primers were used (see Table A-9, page 265, for the primer list). One PCR reaction was performed per primer pair (24X IgH reactions (i.e. 8X3), 4X IgK reaction, 9X IgL reactions). IgG, IgM, IgA, IgK and IgL sequences were amplified (see Figure 2-7 for a schematic). The PCR1 reaction is shown in Table 2-39.

Table 2-39: Ribosome display PCR1 reagents and thermal cycle. See Table A-9 for primers. See Figure 2-7 for schematic.

Reagent	Quantity
Template cDNA (Table 2-38)	1 µl
Forward primer (10 µM)	2.5 µl
Reverse primer (10 µM)	2.5 µl
H_2O	6.5 µl
NEBNext Q5 Hot Start HiFi PCR Master Mix[62]	12.5 µl
Total	25 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X32
Annealing	59-72°C (see Table A-9)	30 s	
Extension	72°C	30 s	
Final Extension	72°C	5 min	X1
Hold	4°C	∞	

25 µl of each PCR1 product was mixed with 5 µl 6X gel loading dye (B7024S, New England Biolabs) and the whole 30 µl was run on a 1.3% (w/v) agarose gel (TAE buffer) stained with SYBR Safe (S33102, Invitrogen) at 90 V for 60 min. The bands were excised using a scalpel. Expected band sizes were as follows: IgH (~387-462 bp), IgK (~650 bp) and IgL (~650 bp). The DNA was purified using the QIAquick Gel Extraction Kit (28704, Qiagen) following the manufacturer's instructions. DNA was eluted in 30 µl H_2O and the concentration was measured using a NanoDrop (Thermo Scientific).

[62] NEBNext Q5 Hot Start HiFi PCR Master Mix (M0543S, NEB)

2.14.3 Addition of Linker Regions (PCR2)

To amplify and add linker sequences to the PCR1 products, PCR2 primers were used as described in Table 2-40. As the heavy chain sequences were amplified using a single reverse primer in PCR2, seven master mixes of heavy chain PCR1 products were made as described in Table 2-40 by adding equal volumes of each heavy chain PCR1 product. The heavy chain master mixes were used as the PCR starting material in PCR2. There was a 1:1 pairing of light chain PCR1 and PCR2 products so no master mixes were necessary.

Table 2-40: Ribosome display library PCR1 and PCR2 primer pairs. PCR2 primer sequences are given in Table A-10 (page 266).

PCR1		PCR Product	PCR2		
Forward	Reverse	Master Mix	Forward	Reverse	Chain
VH1	GCH1link				Heavy Chain
VH1	MCH1link	1	VH1A	GMACH1link	
VH1	ACH1link				
VH1257	GCH1link				
VH1257	MCH1link	2	VH1257A	GMACH1link	
VH1257	ACH1link				
VH2	GCH1link				
VH2	MCH1link	3	VH2A	GMACH1link	
VH2	ACH1link				
VH3A	GCH1link				
VH3A	MCH1link				
VH3A	ACH1link	4	VH3AA	GMACH1link	
VH3B	GCH1link				
VH3B	MCH1link				
VH3B	ACH1link				
VH4	GCH1link				
VH4	MCH1link	5	VH4A	GMACH1link	
VH4	ACH1link				
VH4DP63	GCH1link				
VH4DP63	MCH1link	6	VH4DP63A	GMACH1link	
VH4DP63	ACH1link				
VH6	GCH1link				
VH6	MCH1link	7	VH6A	GMACH1link	
VH6	ACH1link				
VK1	KC		VK1link	KCNotI	Kappa Chain
VK2346	KC		VK2346link	KCNotI	
VK36	KC		VK36link	KCNotI	
VK5	KC		VK5link	KCNotI	
VL1	LC		VL1link	LCNotI	Lambda Chain
VL1459	LC		VL1459link	LCNotI	
VL15910	LC		VL15910link	LCNotI	
VL2	LC		VL2link	LCNotI	
VL3a	LC		VL3alink	LCNotI	
VL3b	LC		VL3blink	LCNotI	
VL3DLP16	LC		VL3DLP16link	LCNotI	
VL6	LC		VL6link	LCNotI	
VL78	LC		VL78link	LCNotI	

The PCR conditions are described in Table 2-41. The heavy chain reverse primer and the light chain forward primers had complementary overhangs, thus allowing them to join during the overlap extension step as described in section 2.14.4.

Table 2-41: Ribosome display PCR2 reagents and thermal cycle. The quantities given are for each PCR reaction. In the heavy chain reaction, 3 µl of IgH PCR1 master mix is used, in the light chain reactions, 1 µl of light chain PCR1 product is used. See Table 2-40 for primer pairs and Table A-10 for primer sequences. See Figure 2-7 for schematic.

Reagent		IgH reaction	IgK/IgL reaction
PCR1 product		3 µl	1 µl
Forward primer (10 µM)		2.5 µl	2.5 µl
Reverse primer (10 µM)		2.5 µl	2.5 µl
H_2O		4.5 µl	6.5 µl
NEBNext Q5 Hot Start HiFi PCR Master Mix[63]		12.5 µl	12.5 µl
	Total	25 µl	25 µl
Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X32
Annealing	59-72°C (see Table A-10)	30 s	
Extension	72°C	30 s	
Final Extension	72°C	5 min	X1
Hold	4°C	∞	

The PCR2 products were run on a 1.3% agarose gel, the bands were excised and DNA was purified and quantified exactly as stated for PCR1 (section 2.14.2).

[63] NEBNext Q5 Hot Start HiFi PCR Master Mix (M0543S, NEB)

2.14.4 Overlap Extension to Link Heavy and Light Chain Variable Regions (PCR3)

3 µl of each PCR2 product was pooled to create three PCR2 master mixes: IgH, IgK and IgL. These master mixes were then used in an overlap extension PCR (PCR3) to randomly link heavy chains with kappa or lambda light chains.

Heavy and light chains were joined using an assembly PCR as described in Table 2-42.

Table 2-42: Ribosome display assembly PCR reagents and thermal cycle. Equal quantities of heavy and light chain master mixes were used to ensure an approximate 1:1 ratio of heavy and light chain sequences.

Reagent	Quantity
IgH PCR2 master mix	60 ng
IgK/IgL PCR2 master mix	60 ng
H_2O	to 25 µl
NEBNext Q5 Hot Start HiFi PCR Master Mix[64]	12.5 µl
Total	25 µl

Thermal Cycle Step	Temperature	Time	Cycles
Denaturation	98°C	1 min	X10
Extension	72°C	2 min	
Hold	4°C	∞	

The kappa and lambda assembly PCR products were then used in 24X kappa PCR3 reactions and 24X lambda PCR3 reactions per donor (48X PCR3 reactions per donor in total) as described in Table 2-43.

Table 2-43: Ribosome display PCR3 reagents and thermal cycle. See Table A-11 for primers. See Figure 2-7 for schematic.

Reagent	Quantity	
	X1	X24
Assembly PCR product	0.5 µl	13 µl
RDT7AB primer (10 µM)	1.25 µl	32.5 µl
ASKANG primer (10 µM)	1.25 µl	32.5 µl
H_2O	9 µl	234 µl
DMSO	0.5 µl	13 µl
NEBNext Q5 Hot Start HiFi PCR Master Mix[65]	12.5 µl	325 µl
Total	25 µl	650 µl

Thermal Cycle Step	Temperature	Time	Cycles
Initial denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X32
Annealing	67°C	30 s	
Extension	72°C	90 s	
Final Extension	72°C	5 min	X1
Hold	4°C	∞	

[64] NEBNext Q5 Hot Start HiFi PCR Master Mix (M0543S, NEB)
[65] NEBNext Q5 Hot Start HiFi PCR Master Mix (M0543S, NEB)

Each group of 24 PCR3 reaction products was pooled and purified using the Wizard SV Gel and PCR Clean-Up System (A9281, Promega) as detailed in section 2.2.4 (page 88) except the PCR3 product was eluted in 60 µl H_2O (40 µl + 20 µl). The entire elutedPCR3 product was mixed with the appropriate volume of 6X gel loading dye (B7024S, New England Biolabs) and was run on a 0.8% (w/v) agarose gel (TAE buffer) stained with SYBR Safe (S33102, Invitrogen) at 90 V for 30 min.

Bands of the correct size (1.2 kb) were excised using a scalpel and the DNA was purified using the Wizard SV Gel and PCR Clean-Up System (A9281, Promega). 10 µl Membrane Binding Solution was added per 10 mg of gel. This was then heated at 56°C until the gel melted. The rest of the protocol was as detailed in section 2.2.4. The DNA was eluted in 30 µl H_2O (20 µl + 10 µl) and quantified using NanoDrop.

This protocol resulted two libraries per donor: one using kappa light chains, the other using lambda light chains.

2.15 Ribosome Display: Panning Selection against EBOV GP

Panning selection of the ribosome display libraries (section 2.14) against Zaire Ebola virus (EBOV) GP (section 2.17) was used to isolate EBOV GP-binding scFv from the repertoires of the three Caucasian Ebola survivors (section 2.2.1.1). A schematic of the ribosome display protocol is shown in Figure 2-8.

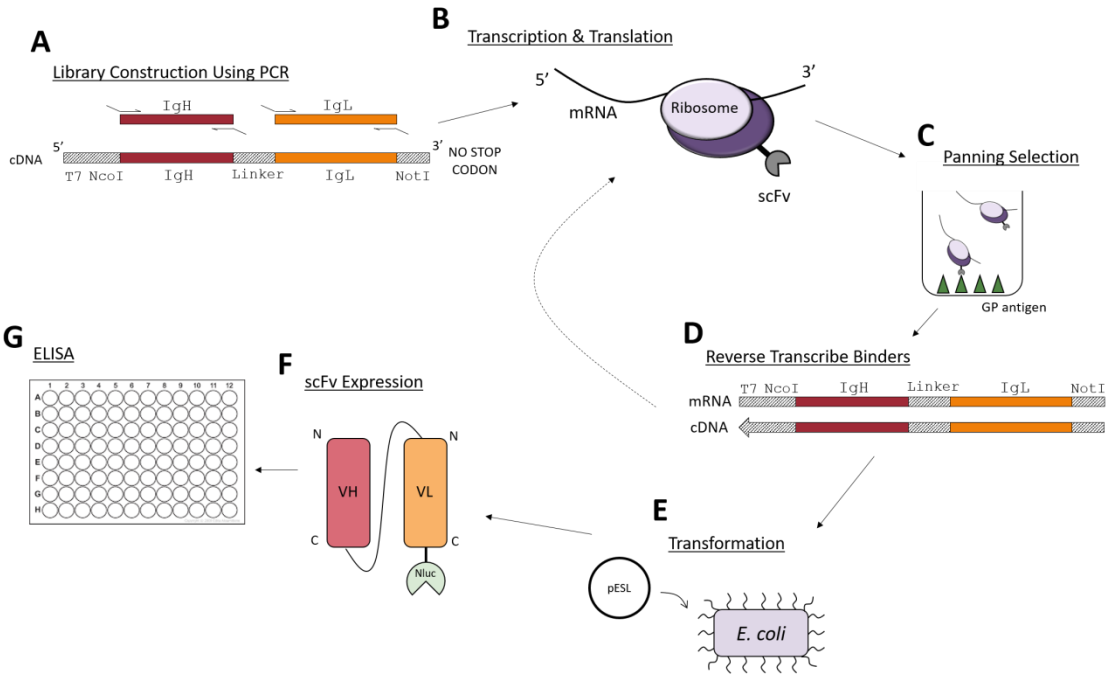


Figure 2-8: Ribosome display method used to isolate Ebola GP-binders. **A)** The ribosome display libraries were built from functional immunoglobulin heavy and light chain variable region sequences which were reverse transcribed from mRNA, amplified by PCR and joined at random using overlap extension PCR (see section 2.14, page 118). The ribosome display libraries lack stop codons. **B)** The ribosome display libraries were transcribed and translated and the scFv mRNA and protein remained connected via the ribosome (due to the lack of a stop codon). **C)** The scFv-ribosome complexes were incubated with the EBOV GP antigen. The scFv which did not bind were washed away. **D)** The scFv-ribosome complexes which did bind the EBOV GP were recovered and the mRNA was reverse-transcribed. **E)** The cDNA was PCR-amplified, inserted into the pESL expression vector and transformed into *E. coli*. **F)** The *E. coli* were plated out and cultured overnight. Individual colonies were picked and cultured in liquid media and scFv conjugated to NanoLuc (Nluc) luciferase were expressed. **G)** Specificity of scFv was confirmed using ELISA. Binding was detected using a luciferase assay.

2.15.1 Panning Selection against GP Antigen

Nunc Maxisorp ELISA plates were coated with 60 µl of 20 µg/ml EBOV GP antigen (section 2.17, page 132) in molecular-grade PBS (10010023, Gibco) (one well per kappa library, one well per lambda library). The plate was incubated overnight at 4°C. The wells were washed once with 300 µl molecular-grade PBS. The wells were blocked with 200 µl of 100 µg/ml purified BSA (B9001S, New England Biolabs) in molecular-grade PBS and incubated for 1 h at room temperature.

To transcribe and translate the ribosome display libraries (section 2.14.4), the reaction shown in Table 2-44 was set up for each kappa and lambda library per donor. The reagents were mixed by pipetting and then centrifuged briefly. The reaction was incubated at 30°C for 90 min.

Table 2-44: Coupled transcription/translation reaction for scFv expression.

Kappa Library		Lambda Library	
Reagents	Quantities	Reagents	Quantities
TNT T7 Quick Master Mix[66]	40 µl	TNT T7 Quick Master Mix[66]	40 µl
Methionine (1 mM)[66]	1 µl	Methionine (1 mM)[66]	1 µl
T7 TnT PCR Enhancer[66]	1 µl	T7 TnT PCR Enhancer[66]	1 µl
VH+VK PCR3 product[67]	≤8 µl	VH+VL PCR3 product[67]	≤8 µl
dH_2O	To 50 µl	dH2O	To 50 µl
Total	50 µl	Total	50 µl

4 µl RNaseOUT recombinant ribonuclease inhibitor (40 U/µl) (10777-019, Invitrogen) was added to each transcription/translation reaction (54 µl total per tube) and the tubes were incubated on ice until use.

Blocking buffer was washed out of the antigen-coated wells by washing twice with 300 µl ice-cold molecular grade PBS and the 54 µl transcription/translation reaction products were applied to the antigen-coated wells. The mixture was incubated on a rocker at 4°C for 1 h. The wells were washed three times with 300 µl ice-cold molecular-grade PBS to remove unbound scFv-mRNA-ribosome complexes.

[66] T7 TnT Quick Coupled Transcription/Translation System (L1170, Promega)
[67] PCR3 product (section 2.14.4), ideally 1 µg DNA in ≤5 µl

The GP-bound scFv-mRNA-ribosome complexes were extracted from the wells using the RNeasy Micro Kit (74004, Qiagen) following the manufacturers quick start protocol with the following notes:

- 7 µl β-mercaptoethanol was added per 350 µl Buffer RLT, this was then pipetted up and down in the well to remove scFv-mRNA-ribosome complexes and then transferred to a clean 1.5 ml microcentrifuge tube
- 5 µl of 4 ng/µl carrier (poly-A) RNA was added to each tube
- On-column DNaseI digestion was performed
- The mRNA was eluted in 14 µl H_2O

Reverse transcription primers were annealed to the mRNA by mixing mRNA, primers and dNTPs as described in Table 2-45. The mixture was incubated at 70°C for 5 min, then put on ice. The reverse transcription reaction was performed as described in Table 2-46. The reverse-transcribed sequences were amplified using PCR4 as described in Table 2-47.

Table 2-45: Annealing reverse transcription primers to rescued mRNA. HuκF and HuλF primers are used for the kappa and lambda libraries respectively. See Table A-12 for primer sequences.

Reagent	Quantity
mRNA from panning selection (section 2.15.1)	4 µl
HuκF or HuλF primer (10 µM)	5 µl
dNTPs (10 mM each)	1 µl
Total	10 µl

Table 2-46: Reverse transcription reaction reagents and thermal cycle.

Reagent	Quantity
5X Reaction Buffer[68]	4 µl
RNaseOUT (40 U/µl)[69]	0.5 µl
DTT (0.1 M)[68]	2 µl
Nuclease-free H_2O	2.5 µl
Reverse Transcriptase Enzyme[68]	1 µl
Reaction from Table 2-45	10 µl
Total	20 µl

Thermal Cycle Step	Temperature	Time	Cycles
Annealing	25°C	7 min, 30 s	X1
Extension	42°C	60 min	
Enzyme Inactivation	95°C	5 min	
Hold	4°C	∞	

[68] Protoscript II Reverse Transcriptase Kit (M0368, New England Biolabs)
[69] RNaseOUT recombinant ribonuclease inhibitor (40 units/ul) (10777-019, Invitrogen)

Table 2-47: Ribosome display PCR4 reagents and thermal cycle. See Table A-13 for primer sequences.

Reagent		Quantity	
		X1	X8
Reverse transcribed cDNA (Table 2-46)		1 µl	8 µl
RDT7AB_short primer (10 µM)		1.25 µl	10 µl
NotK or NotL primer (10 µM)		1.25 µl	10 µl
Nuclease-free H_2O		9 µl	72 µl
Q5 Master Mix[70]		12.5 µl	100 µl
	Total	25 µl	200 µl
Thermal Cycle Step	Temperature	Time	Cycles
Initial Denaturation	98°C	1 min	X1
Denaturation	98°C	10 s	X32
Annealing	68°C	30 s	
Extension	72°C	1 min	
Final Extension	72°C	5 min	X1
Hold	4°C	∞	

The PCR4 products were pooled and purified in 30 µl H_2O using the Wizard SV Gel and PCR Clean-Up System (A9281, Promega) as detailed in section 2.2.4.

The pooled PCR4 products were run on a 0.8% (w/v) agarose gel at 100 V for 20 min. The bands (760-800 bp) were excised using a scalpel and purified using the QIAquick Gel Extraction Kit (28704, Qiagen) in 30 µl H_2O. The concentration was measured using NanoDrop.

The PCR4 output sequences are shorter than the PCR3 input sequences because the reverse transcription primers bind further upstream in the light chain constant region. This is to increase yield of rescued PCR product because the 3' end of the mRNA may have been damaged when the ribosome was removed.

2.15.2 Insertion of scFv Sequences into pESL Plasmids

The pESL plasmid (see Figure A-2, page 271, for plasmid map) and PCR4 product were digested using the reaction shown in Table 2-48. The reaction was incubated overnight at 37°C, then 80°C for 20 min to inactivate the enzymes.

[70] NEBNext Q5 Hot Start HiFi PCR Master Mix (M0543S, New England Biolabs)

Table 2-48: Restriction digest of PCR4 product and pESL plasmid.

pESL plasmid digestion		PCR4 insert digestion	
Reagent	Quantity	Reagent	Quantity
pESL plasmid	1 µg	PCR4 product	0.5 µg
CutSmart Buffer (10X)[71]	5 µl	CutSmart Buffer (10X)[71]	2.5 µl
NcoI-HF (20,000 U/ml)[72]	1 µl	NcoI-HF (20,000 U/ml)[72]	0.5 µl
NotI-HF (20,000 U/ml)[73]	1 µl	NotI-HF (20,000 U/ml)[73]	0.5 µl
Nuclease-free H_2O	To 50 µl	Nuclease-free H_2O	To 25 µl
Total	50 µl	Total	25 µl

The digested DNA was run on a 0.8% (w/v) agarose gel at 90 V. The bands were excised using a scalpel and the DNA was purified using QIAquick Gel Extraction Kit (28704, Qiagen). The pESL plasmid was eluted in 50 µl H_2O and the PCR4 product was eluted in 30 µl H_2O. The concentration of eluted DNA was measured using NanoDrop.

The PCR4 insert library was ligated into pESL plasmids using a 1:3 plasmid to insert molar ratio as set out in Table 2-49. The reaction was incubated at room temperature for 1 h.

Table 2-49: Ligation of PCR4 product into pESL plamid. The control ligation lacks insert DNA and gauges the background due to incomplete digestion.

Reagent	Quantity	
	Ligation	Control
T4 DNA Ligase Buffer (10X)[74]	2 µl	2 µl
Digested pESL (Table 2-48)	20 ng	20 ng
Digested PCR4 product (Table 2-48)	8.5 ng	-
T4 DNA Ligase[74]	1 µl (400 U)	1 µl (400 U)
Nuclease-free H_2O	To 20 µl	To 20 µl
Total	20 µl	20 µl

5 µl of ligation mix (Table 2-49) was transformed into NEB 5-alpha Competent *E. coli* (C2987R, New England Biolabs) following the manufacturer's instructions. The transformed *E. coli* were plated on LB agar (Miller)[75] supplemented with 50 µg/ml kanamycin (BP9065, Fisher Scientific) and incubated overnight at 37°C.

94 colonies were picked per kappa and lambda library per donor and colony PCR was used to determine presence of PCR4 insert in each pESL vector. Colony PCR was performed as described in Table 2-50 and the product was run on 0.8% agarose gel. pESL containing insert

[71] CutSmart Buffer (10X) (B7204S, New England Biolabs)
[72] NcoI-HF (20,000 units/ml) (R3193S, New England Biolabs)
[73] NotI-HF (20,000 units/ml) (R3189S, New England Biolabs)
[74] T4 DNA Ligase (M0202S, New England Biolabs)
[75] Tryptone 10 g/L, Yeast extract 5 g/L, Sodium Chloride 10 g/L (Fisher BioReagents Microbiology Media)

has a band of approximately 1600 bp. Empty pESL plasmid has a band of 980 bp. Only colonies with inserts of the correct size were tested in an ELISA against EBOV GP (section 2.16).

Table 2-50: Colony PCR to find inserts of the correct size. See Table A-14 for primer sequences.

Reagent		Quantity	
		X1	X110
T7F primer (10 µM)		0.5 µl	55 µl
T7R primer (10 µM)		0.5 µl	55 µl
One*Taq* Quick-Load 2X Master Mix[76]		12.5 µl	1375 µl
Nuclease-free H_2O		11.5 µl	1265 µl
	Total	25 µl	2750 µl
Thermal Cycle Step	Temperature	Time	Cycles
Initial Denaturation	98°C	30 s	X1
Denaturation	98°C	10 s	X32
Annealing	44°C	30 s	
Extension	72°C	1 min	
Final Extension	72°C	5 min	X1
Hold	4°C	∞	

[76] One*Taq* Quick-Load 2X Master Mix (M0486L, New England Biolabs)

2.16 Ribosome Display: Screening scFv for GP Binding (ELISA)

Nunc FluoroF96 White MaxiSorp ELISA plates were coated with the desired concentration of glycoprotein (GP) (section 2.17, page 132) diluted in PBS and incubated at 4°C overnight. The wells were washed three times with 0.05% (v/v) Tween-20 in PBS, blocked with 200 µl blocking buffer (3% (w/v) skimmed milk powder in PBS) and incubated at room temperature for 1 h. The wells were washed three times with 0.05% (v/v) Tween-20 in PBS.

Individual scFv-expressing *E. coli* colonies were picked and grown in 1 ml LB broth (Miller)[77] containing 50 µg/ml kanamycin (BP9065, Fisher Scientific) at 26°C, 200 rpm for 22 h. The bacteria were pelleted by centrifugation at 4500 rpm for 10 min. 50 µl of clarified supernatant was applied to each GP-coated ELISA well and incubated at room temperature for 2 h. The wells were washed three times with 0.05% (v/v) Tween-20 in PBS. 50 µl Nano-Glo solution (Table 2-51) was applied to each well and incubated at room temperature for 10 min. Luminescence was read on a Victor X Light plate reader (Perkin Elmer) (see Figure 2-9).

Colonies expressing binders were sequenced using Sanger sequencing (Eurofins Genomics, Germany) using the T7F and pESL_SeqR primers (see Table A-14 for primer sequences).

Table 2-51: Reagents to make up Nano-Glo substrate mixture.

Reagent	Quantity (X1 reaction)
PBS	25 µl
Nano-Glo Luciferase Assay Buffer[78]	25 µl
Nano-Glo Luciferase Assay Substrate[78]	0.5 µl
Total	50.5 µl

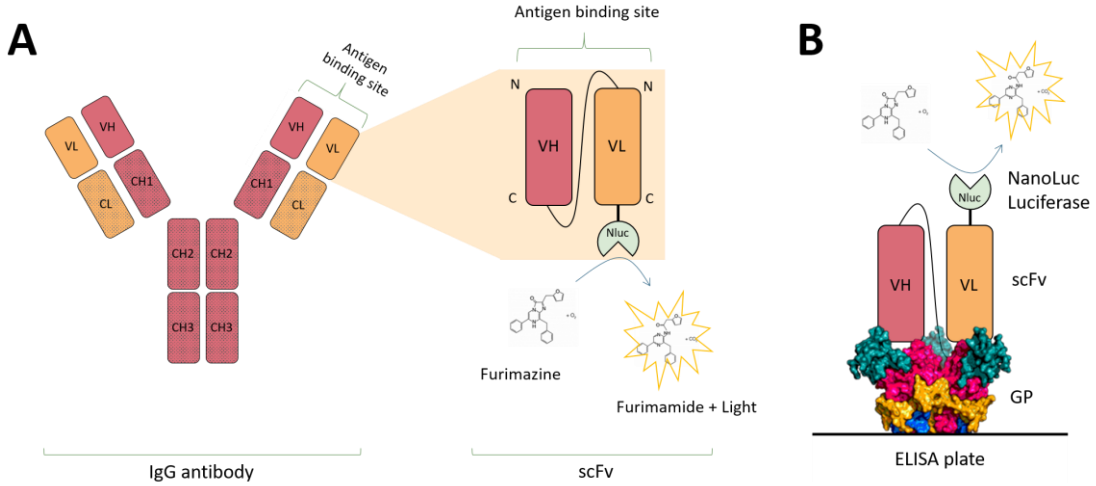


Figure 2-9: Ribosome display scFv conjugated to NanoLuc luciferase. **A)** scFv are the immunoglobulin variable region held together by a linker instead of the antibody constant region. **B)** A secondary antibody was not required in the ELISA as the scFv were conjugated to NanoLuc luciferase, which breaks down Furimazine to create luminescence.

[77] Tryptone 10 g/L, Yeast extract 5 g/L, Sodium Chloride 10 g/L (Fisher BioReagents Microbiology Media)
[78] Nano-Glo Luciferase Assay System (N1110, Promega)

2.17 Expression of *Filoviridae* Glycoprotein Ectodomains

2.17.1 Glycoprotein Ectodomain Sequences[79]

The nucleotide sequences encoding the glycoprotein (GP) ectodomains of Zaire Ebola virus (Mayinga strain)[80] and Marburg virus (Ravn strain)[80] (Table 2-52) were cloned into the pHL-sec expression plasmid (Aricescu et al., 2006) using restriction sites EcoRI and KpnI (see Figure A-3, page 271, for plasmid map). The ectodomain sequence was inserted upstream of a His-tag.

The plasmids were cloned into NEB 10-beta *E. coli* (C3019H, New England Biolabs) and the bacteria were grown on LB agar/broth (Miller)[81] supplemented with 100 µg/ml ampicillin.

Plasmids were purified using NucleoBond Xtra Maxi prep kit (Macherey-Nagel) and eluted in H_2O.

Table 2-52: *Filoviridae* glycoprotein (GP) ectodomains expressed. GenBank accession number and the amino acid residues of expressed Ebola and Marburg virus GP ectodomain are shown. The cytoplasmic domain and transmembrane region of each GP was excluded.

Filoviridae species	Strain	GenBank ID	Residues	Reference
Zaire Ebola virus (*Zaire ebolavirus*)	Mayinga	U23187.1	1-632	Sanchez et al. (1996)
Marburg virus (*Marburg marburgvirus*)	Ravn	ABE27071.1	1-642	Towner et al. (2006)

2.17.2 Glycoprotein Expression[82]

FreeStyle 293-F cells (Gibco) were seeded in FreeStyle 293 Expression Medium at a density of 0.7 X 10^6 cells/ml. The cells were transfected the following day with plasmid DNA at a concentration of 2 µg/ml in a 1:6 ratio of DNA:PEI Max[83].

To transfect 1 L of FreeStyle 293-F cells, 2 mg of plasmid DNA was diluted in Opti-MEM Reduced Serum Media (31985062, Gibco) to a final volume of 50 ml and 12 ml PEI Max was diluted in Opti-MEM to a final volume of 50 ml. The two solutions were mixed together (100 ml total) and incubated for 30 min at room temperature before adding to the 1 l of cell culture.

Cells were incubated at 37°C, 7% CO_2, 135 rpm for 5 days.

[79] The work detailed in section 2.17.1 was conducted by Audrey Page (KCL)
[80] Nucleotide sequences encoding the GP ectodomains were generously shared by Stuart Neil (KCL)
[81] Tryptone 10 g/L, Yeast extract 5 g/L, Sodium Chloride 10 g/L (Fisher BioReagents Microbiology Media)
[82] Zaire Ebola virus GP was expressed by Audrey Page (KCL) and Catherine Townsend (author), Marburg (Ravn) virus GP was expressed by Catherine Townsend (author)
[83] Polyethylenimine Hydrochloride, Linear (MW 4,000) (24885, Polysciences Inc.) at 1 mg/ml in H_2O

2.17.3 Glycoprotein Purification[84]

FreeStyle 293-F cell culture supernatant was clarified by centrifugation at 3000 rpm for 10 min at 4°C. The clarified supernatant was passed through a 0.22 µm filter. 2 M NaCl[85] and 2 M Imidazole[85] were added to the supernatant to a final concentration of 0.5 M NaCl, 30 mM imidazole, pH 7.4. The supernatant was kept on ice throughout the protocol.

The His-tagged GP was purified by gravity flow separation using an Econo-Column 1.5 X 5 cm Chromatography Column (Bio-Rad) manually packed with 2 ml Ni Sepharose 6 Fast Flow slurry (17-5318-06, GE Healthcare). The slurry was not allowed to dry out at any point. The column was washed with five volumes of 0.22 µm filtered dH_2O. The column was then equilibrated with 10 volumes of Binding Buffer[85]. The supernatant was loaded on to the column and filtered through the slurry by gravity. The column was then washed with 20 ml of Washing Buffer[85]. The His-tagged GP was eluted in 5 ml Elution Buffer (300 mM)[85] then 5 ml Elution Buffer (500 mM)[85]. The eluent was collected in 1 ml fractions.

A western blot was used to determine which eluent fractions contained protein. 10 µl of each 1 ml fraction was mixed with 2 µl 6X Loading Buffer[85] and run on a 0.75 mm 10% SDS-PAGE gel[85] in SDS Running Buffer[85] at 100 V for 1 h. The protein was transferred to a membrane in Transfer Buffer[85] at 100 V for 1 h. The membrane was blocked for 1 h at room temperature in blocking buffer (5% (w/v) skim milk powder, 0.05% (v/v) Tween-20 in PBS). The membrane was incubated for 1 h at room temperature in 6X-His Tag Monoclonal Antibody (4A12E4) (37-2900, Invitrogen) diluted 1/5000 in blocking buffer. The membrane was washed four times in 0.05% (v/v) Tween-20 in PBS and incubated for 30 min at room temperature in IRDye 680RD Goat (polyclonal) Anti-Mouse IgG (H+L) (926-68070, LI-COR) diluted 1/5000 in blocking buffer. The membrane was washed four times in 0.05% (v/v) Tween-20 in PBS and then kept in PBS until it was imaged. The membrane was imaged at a wavelength of 700 nm.

The fractions containing protein were pooled, loaded into an Amicon Ultra-15 Centrifugal Filter Unit (UFC900308, Merck Millipore) and centrifuged at 3000 rpm for 30 min at 4°C. The flow-through was discarded and 15 ml of PBS was loaded on to the column and centrifuged under the same conditions. This step was repeated once more. Approximately 200 µl of PBS containing protein was collected in the filter. The protein concentration was measured by NanoDrop and the purity was confirmed using an SDS-PAGE gelSDS-PAGE gel[85] and Coomassie[85] stain.

[84] Zaire Ebola virus GP was purified by Audrey Page (KCL) and Catherine Townsend (author), Marburg (Ravn) virus GP was purified by Catherine Townsend (author)

[85] See pages 272 to 273 for buffer and gel recipes

2.17.4 Deglycosylation of Glycoproteins

To deglycosylate EBOV GP, RAVV GP and HIV gp120[86], 20 µg of glycoprotein was diluted to 42 µl with PBS and incubated with 4 µl of Endo H (4.7 mg/ml)[87] and 4 µl PNGase F (15 mg/ml)[87] (final glycoprotein concentration of 400 µg/ml in a volume of 50 µl). The reaction was incubated overnight at 37°C. As a control, 20 µg of glycoprotein was diluted to 50 µl in PBS only (no enzymes) and incubated at 37°C overnight.

The success of the deglycosylation was checked by running a non-reducing 10% SDS-PAGE. The deglycosylated protein should run at a lower molecular weight than the complete protein.

Additionally, an ELISA was used to check that PHA-L lectin bound less to the deglycosylated protein than the complete protein. 50 µl of complete and deglycosylated protein at 5 µg/ml in PBS was coated in triplicate onto a Nunc Maxisorp plate and incubated overnight. The plate was washed once with 0.05% (v/v) Tween-20 in PBS. The plate was incubated for 1 h at room temperature with 200 µl blocking buffer (3% (w/v) skim milk powder, 0.05% (v/v) Tween-20 in PBS). The plate was washed three times as above. 50 µl 5 µg/ml Biotinylated *Phaseolus vulgaris* leucoagglutinin (PHA-L) (B-1115, Vector Laboratories) in PBS was added to each well and incubated for 1 h at room temperature. The plate was washed three times as described above. 50 µl alkaline phosphatase-conjugated streptavidin (SA1008, Invitrogen) was added to each well at a 1/10,000 dilution in blocking buffer (as above) and incubated for 1 h at room temperature. The plate was washed three times as above and developed using SIGMA*FAST* p-Nitrophenyl phosphate Tablets (N1891, Sigma Aldrich) as described by the manufacturer. The plates were read at a wavelength of 405 nm.

[86] HIV gp120 was expressed and purified by Luke Granger (KCL)

[87] Endo H and PNGase F were expressed and purified by Isabella Hüttner (KCL) and Elda Iljazi (KCL)

Chapter 3

Tolerance, Autoreactivity and Polyreactivity

Chapter 3: Tolerance, Autoreactivity and Polyreactivity

3.1 Introduction

Autoreactive antibodies are highly undesirable in the peripheral repertoire as they pose the risk of inducing autoimmune disease. Polyreactivity is closely linked to autoreactivity as antibodies which bind multiple antigens are more likely to bind self-antigen. The ability to identify antibodies which are likely to be autoreactive and/or polyreactive based on their physicochemical or structural properties would be extremely valuable in the diagnosis of autoimmune disease and also in the pharmaceutical industry where it is imperative that therapeutic antibodies are highly specific to their target and not autoreactive.

As autoreactive antibodies are potentially dangerous if released into the peripheral blood, it is thought that B cells bearing autoreactive antibodies are removed from the B cell repertoire at two tolerance checkpoints: central tolerance and peripheral tolerance. As has been described in detail in sections 1.4 and 1.5 (page 37 and 39 respectively), it is thought that central tolerance acts on the immature B cell repertoire in the bone marrow to remove the majority of autoreactive B cells before they enter the peripheral blood (Wardemann et al., 2003) and peripheral tolerance removes autoreactive B cells in the peripheral blood which escaped central tolerance. Relatively few characteristics associated with autoreactivity or polyreactivity have been identified; they tend to be long, hydrophobic and positively charged CDR-H3 loops. IGHV4-34 has also been associated with numerous autoimmune diseases (Tipton et al., 2015, Doorenspleet et al., 2014, Baranzini et al., 1999).

With high-throughput sequencing of B cell repertoires now possible, we hypothesised that by sequencing B cell repertoires isolated from the bone marrow and peripheral blood, we could identify immunoglobulin variable region properties which may be associated with autoreactivity by looking for properties which are less prevalent in the peripheral blood repertoires compared to the bone marrow repertoires. Variable region properties associated with autoreactivity should have been selected against at the tolerance checkpoints.

The dataset used for this analysis was the B cell Development dataset. The B cell Development dataset is a high-throughput sequencing dataset constructed from immunoglobulin mRNA isolated from bone marrow B cells (pre-B and immature) and peripheral blood B cells (transitional and naïve) isolated from healthy adult donors. The dataset contains information on VDJ gene usage and CDR3 physicochemical properties. For full materials and methods of

how the B cell Development dataset was constructed, see section 2.1 and (page 74) and section 2.4 (page 94).

The results of the analysis of the B cell Development dataset were used to select some immunoglobulin heavy chain variable regions from the B cell Development dataset which possessed properties which may be associated with autoreactivity. These hypotheses were then tested by cloning antibodies containing the selected heavy chains and investigating their autoreactivity using HEp-2 immunocytochemistry. As polyreactivity is commonly associated with autoreactivity, the antibodies were also tested for polyreactivity using a solid-phase ELISA in which the antibodies were tested for binding to a range of autoantigens following the protocol set out in Tiller et al. (2008). This work led to the further investigation of the nature of binding of the antibodies which displayed a polyreactive phenotype in solid-phase ELISA.

The main focus of this chapter is on the heavy chain properties because, firstly, the heavy chain is thought to be the most important region for antibody specificity (particularly the CDR-H3). Secondly, we found few differences in light chain variable region properties between the bone marrow and peripheral blood in the B cell Development dataset (this analysis is presented in section 4.2, page 192).

When selecting heavy chain properties to investigate, emphasis was placed on the difference in frequency between the pre-B/immature repertoire and the naïve repertoire rather than the frequency in the transitional population. This was due to the controversy surrounding the linearity of B cell development from pre-B >> immature >> transitional >> naïve (Martin et al., 2016). This issue is discussed in this chapter (section 3.7.1, page 178).

In summary, the aim of this chapter is to further our understanding of the common characteristics of autoreactive/polyreactive antibodies using the following methods:

1. Analysis of the heavy chain B cell Development dataset to identify variable region properties which increase or decrease in frequency as B cells mature and therefore may be related to autoreactivity
2. Use this data analysis to identify potentially autoreactive heavy chain sequences, clone and express these as IgG1 antibodies and test for autoreactivity using immunocytochemistry against HEp-2 cells
3. Investigate the nature of polyreactivity and the properties of antibodies which show a polyreactive phenotype in a solid-phase ELISA and try to isolate more polyreactive antibodies using phage display

3.2 Changes in Frequency of Heavy Chain Properties as B cells Mature

3.2.1 VDJ Gene Usage in Bone Marrow and Peripheral Blood B cell Repertoires

The frequency of usage of each IGHV, IGHD and IGHJ family and gene was calculated for the pre-B, immature, transitional and naïve cell repertoires in each of the 12 donors in the heavy chain B cell Development dataset. The mean frequency (+/- 1 SD) of each family/gene across the 12 donors was plotted in bar charts (Figure 3-1) and significant differences in family/gene frequency between cell repertoires was calculated using two-way unpaired ANOVA followed by Tukey's multiple comparisons post-test.

This analysis was performed on the full heavy chain dataset which contained 39,637 entries (Table A-19, page 274). However, data for donor 111 transitional cells was excluded from IGHV-D-J family analysis because there were only 19 sequences in this sample, which was too few to accurately calculate family frequency (see section 2.5.1, page 96, for details on the "5X rule"). Further donors were excluded from the IGHV and IGHD gene usage analysis for this reason and the details are shown in Table 3-1.

Table 3-1: Donors excluded from IGHV and IGHD gene usage analysis. Donors excluded from IGHV gene usage analysis are shown in red (<305 entries). Donors excluded from IGHD gene usage are underlined (<130 entries).

Donor ID	Pre-B	Immature	Transitional	Naïve
103	87	364	65	1606
111	620	74	19	1107
118	2149	478	294	647
119	2966	1704	908	367
122	1290	285	343	259
138	2629	76	195	497
140	731	436	340	571
141	722	1485	979	501
146	3832	675	464	245
149	1120	1135	178	146
159	830	1880	502	349
160	89	191	98	3109

The IGHV3 family was significantly less frequently used in the naïve repertoire compared to the pre-B, immature and transitional repertoires (Figure 3-1A). When looking at the individual genes, nine IGHV3 genes changed significantly in frequency between the cell types, with IGHV3-15, IGHV3-30, IGHV3-33 and IGHV3-7 decreasing in frequency as B cells matured (Figure 3-1D). None of the other IGHV families showed significant differences in usage between the cell types. However, two IGHV1 genes (IGHV1-18 and IGHV1-69) did increase in frequency as the B cells matured (Figure 3-1D). IGHV4-34 and IGHV4-39 were significantly less frequent in the transitional repertoire compared to the pre-B and immature repertoires, however the frequency increased slightly in the naïve repertoire. Conversely, the frequency of IGHV3-30-3 increased in the transitional repertoire and decreased again in the naïve repertoire (Figure 3-1D).

The frequency of the IGHD2 family is significantly lower in the transitional and naïve repertoires compared to the pre-B and immature repertoires (Figure 3-1B). This appears to be driven by a reduction in the frequency of IGHD2-15 and IGHD2-2 genes as B cells matured. IGHD3-3, IGHD4-17 and IGHD6-19 were slightly more frequent as the B cells matured (Figure 3-1E) and the frequency of the IGHD3 family was significantly higher in the transitional repertoire, whereas there was no difference in frequency between the other three repertoires (Figure 3-1B).

The IGHJ3 family is slightly more frequent in the naïve repertoire and the usage of IGHJ6 was significantly lower in the naïve repertoire compared to the pre-B and immature repertoires, however IGHJ6 was used significantly more frequently in the transitional repertoire. IGHJ4 and IGHJ5 were used significantly less frequently in the transitional repertoire (Figure 3-1C).

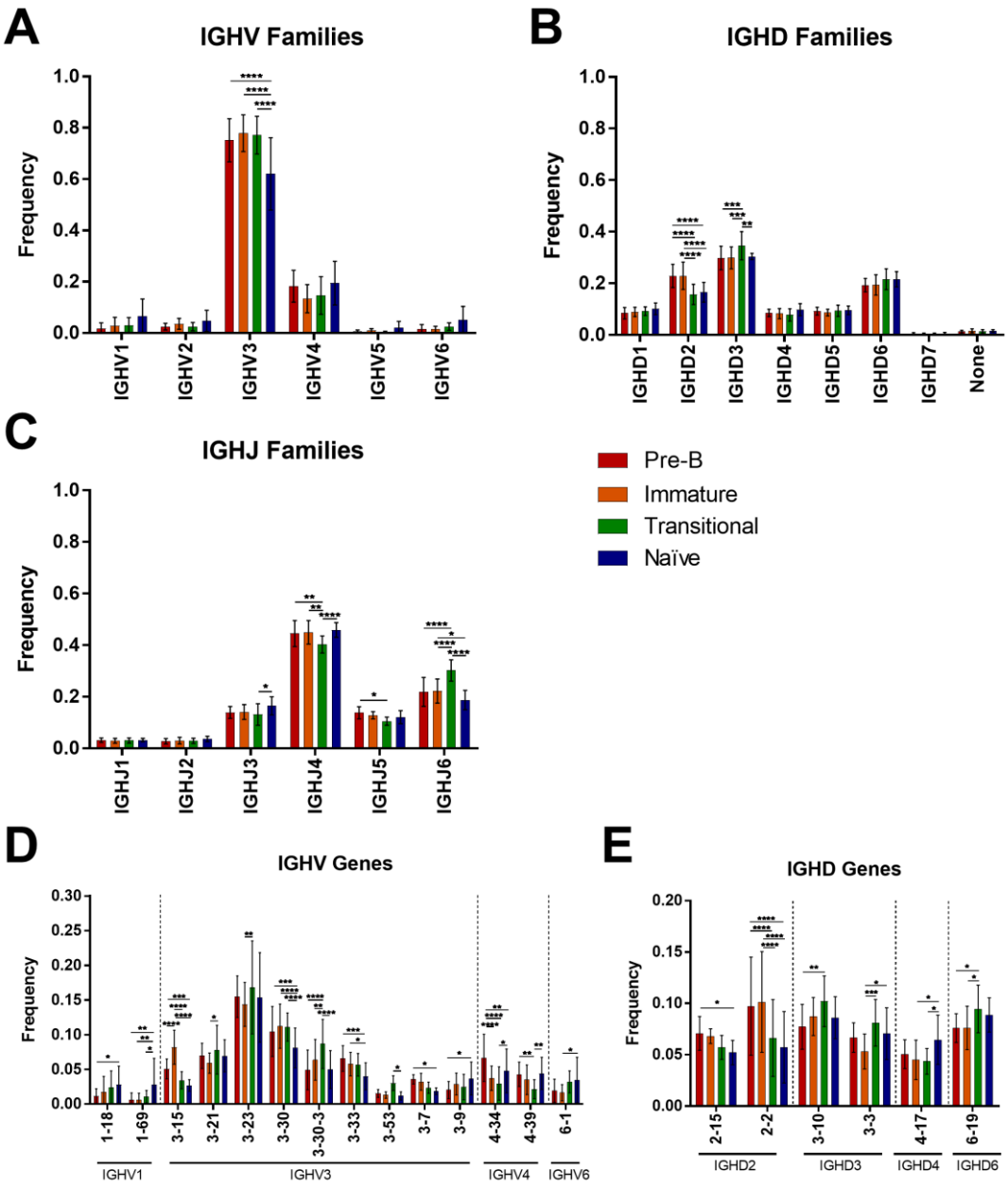


Figure 3-1: BM dataset IGHV/D/J family and gene usage in four cell types. Mean (+/- 1 SD) of the frequency of usage of each (**A**) IGHV, (**B**) IGHD and (**C**) IGHJ family and (**D**) IGHV and (**E**) IGHD genes which showed a significant difference in frequency of usage between cell types across the 12 donors. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

3.2.2 CDR-H3 Properties in Bone Marrow and Peripheral Blood B cell Repertoires

Before performing analysis of the CDR-H3 properties, sequences which may skew the analysis were removed from the heavy chain B cell Development dataset (see section 2.5.2, page 96) resulting in a dataset containing 29,016 variable region sequences (Table A-20, page 274).

The following physicochemical properties were calculated (section 2.4.2, page 94) for each CDR-H3 amino acid sequence (IMGT definition):

1. CDR-H3 length (amino acids)
2. Aliphatic index
3. Boman index
4. Isoelectric point
5. Hydrophobicity (GRAVY index)
6. Instability index

These properties were chosen because, although not orthogonal, they represent a wide range of peptide properties, some of which are known to be important in antibody specificity.

Data for all 12 donors was pooled and the CDR-H3 physicochemical properties of each cell type (pre-B, immature, transitional and naïve) were plotted (Figure 3-2). The Kruskal-Wallis test followed by Dunn's multiple comparisons post-test was used to identify significant differences in CDR-H3 properties between the cell types.

The mean CDR-H3 length was significantly lower in the naïve repertoire compared to the other three repertoires (Figure 3-2A). The naïve repertoire also has a higher Boman index (Figure 3-2C) and both the transitional and naïve repertoires have significantly lower GRAVY indexes than the pre-B and immature repertoires (Figure 3-2E).

The aliphatic index is significantly lower in the transitional repertoire compared to the other three repertoires (Figure 3-2B) and the instability index is significantly lower in the immature repertoire compared to the other three repertoires (Figure 3-2F).

Due to the shape of the isoelectric point (pI) histograms (Figure 3-2D), the differences in the means for each cell type do not reflect the complete picture. The histograms show that very few CDR-H3 have a pI between 5-6 or 7-8.5 in all four cell repertoires. The transitional repertoire has a significantly higher mean pI than the other three repertoires and the histograms show that this is predominantly due to a lower frequency of antibodies with a CDR-H3 pI between 3-5 and a higher frequency of antibodies with a CDR-H3 pI between 8-10.

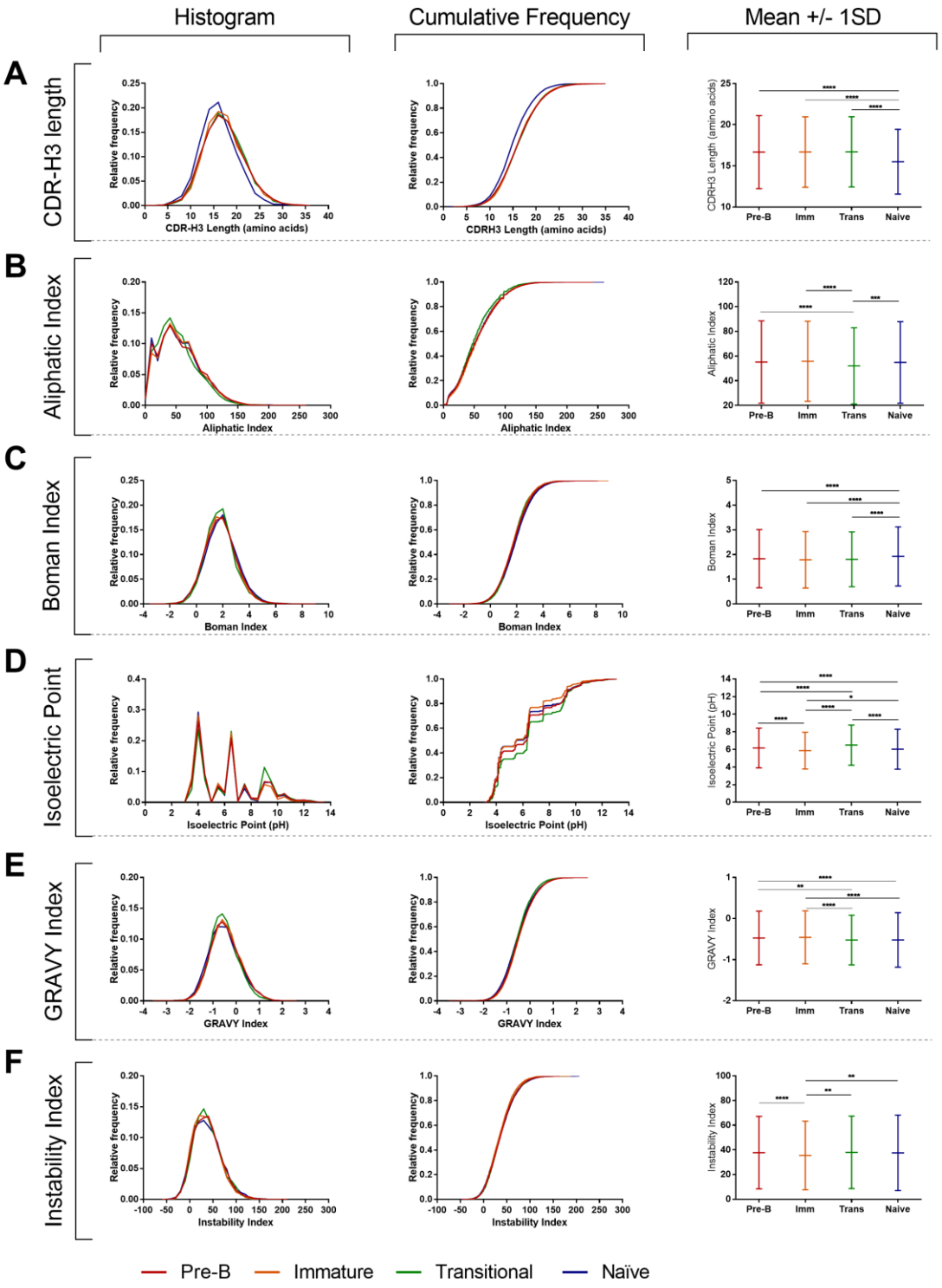


Figure 3-2: BM dataset CDR-H3 properties in four cell types. Histogram, cumulative frequency histogram and mean (+/- 1 SD) for the following six physicochemical properties of the CDR-H3 amino acid sequences: (**A**) CDR-H3 length; (**B**) Aliphatic index; (**C**) Boman index; (**D**) Isoelectric point; (**E**) GRAVY index; (**F**) Instability index. Each set of three graphs shows the same data in different formats. The mean +/- 1 SD graphs are used to see significant differences between cell types, whereas the histograms and cumulative frequency plots show the shape of the data. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

The frequency of the following types of amino acids in the CDR-H3 was measured (see Table 2-20, page 95, for definitions):

1. Tiny
2. Small
3. Aliphatic
4. Aromatic
5. Non-polar
6. Polar
7. Charged
8. Basic
9. Acidic

The mean frequency (+/- 1 SD) of each type of amino acid in the CDR-H3 regions was plotted (Figure 3-3) and the Kruskal-Wallis test followed by Dunn's multiple comparisons post-tests was used to find significant differences between the four cell types.

The frequency of charged and acidic amino acids was significantly higher in the naïve repertoire (Figure 3-3G&I) and the frequency of basic amino acids was significantly higher in transitional and naïve repertoires (Figure 3-3H). Small amino acids were significantly less frequent in the transitional and naïve repertoires (Figure 3-3B). The frequency of all amino acids except tiny and basic was significantly different in the transitional repertoire compared to the other three repertoires (Figure 3 5B, C, D, E, F, G, I).

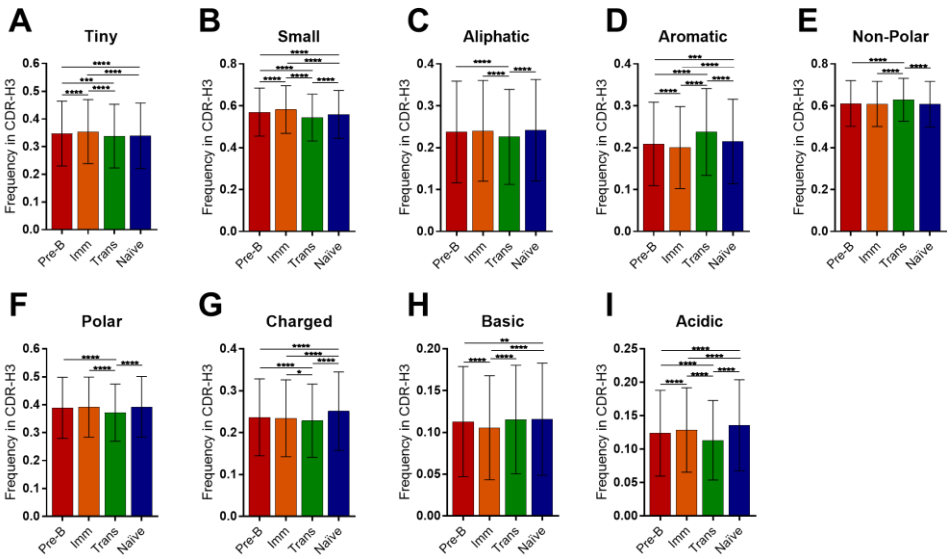


Figure 3-3: Frequency of each amino acid group in the CDR-H3. Mean frequency (+/- 1 SD) of (**A**) Tiny, (**B**) Small, (**C**) Aliphatic, (**D**) Aromatic, (**E**) Non-Polar, (**F**) Polar, (**G**) Charged, (**H**) Basic and (**I**) Acidic amino acids in pre-B, immature, transitional and naïve B cell repertoires. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

3.3 Identification of Potentially Autoreactive Heavy Chains

Six rationales were used to identify potentially autoreactive/polyreactive heavy chains in the B cell Development dataset. These rationales are described in sections 3.3.1 to 3.3.6. The six rationales for heavy chain selection were derived from the data analysis in section 3.2 and also from information in the literature.

As the B cell Development dataset does not contain heavy-light chain paring information, a rationale was required for selecting light chains with which to pair the heavy chains. The 12 heavy chains described in section 3.3.1 were all paired with three public light chains (DeKosky et al., 2015) with very different properties (making 36 antibodies in total). All of the other heavy chains described in sections 3.3.2 to 3.3.6 were paired with the same light chain. Therefore 61 antibodies were expressed in total.

All of the heavy and light chain sequences were identified in the B cell Development dataset and oligonucleotide sequences were ordered from Integrated DNA Technologies (Illinois, USA). The antibodies were expressed as IgG1[88] and purified in PBS. HP-SEC traces are shown in Appendix A.8 (page 280). Five of the antibodies could not be expressed.

3.3.1 Selection of *Forbidden* Heavy Chains

The analysis in section 3.2 showed that IGHV3, IGHD2 and IGHJ6 families were used less frequently either in the naïve repertoire, or in the transitional and naïve repertoires compared to the pre-B and immature repertoires (Figure 3-1A,B&C). Long CDR-H3, hydrophobic CDR-H3 and CDR-H3 with a low Boman index were also used less frequently in the naïve or both the naïve and transitional repertoires (Figure 3-2A,E&C). We hypothesised that this meant that IGHV3, IGHD2, IGHJ6 families and long CDR-H3, hydrophobic CDR-H3 and CDR-H3 with low Boman indexes were selected against at tolerance, and therefore may be associated with autoreactivity/polyreactivity. To identify some heavy chains with these properties, the B cell Development dataset was filtered for sequences which possessed all of the following criteria:

1. IGHV3 family usage
2. IGHD2 family usage
3. IGHJ6 family usage
4. CDR-H3 ≥23 amino acids long
5. CDR-H3 GRAVY index >0.163
6. CDR-H3 Boman index <0.712

[88] All antibodies described in sections 3.3.1 to 3.3.6 were cloned and expressed at MedImmune (Cambridge, United Kingdom) using the method described in section 2.7 (page 99).

The cut-offs for CDR-H3 properties were defined as being either one or two standard deviations from the mean of the naïve B cell repertoire in the six young (<65 years old) donors (122, 159, 138, 118, 146, 103) (Figure 3-4).

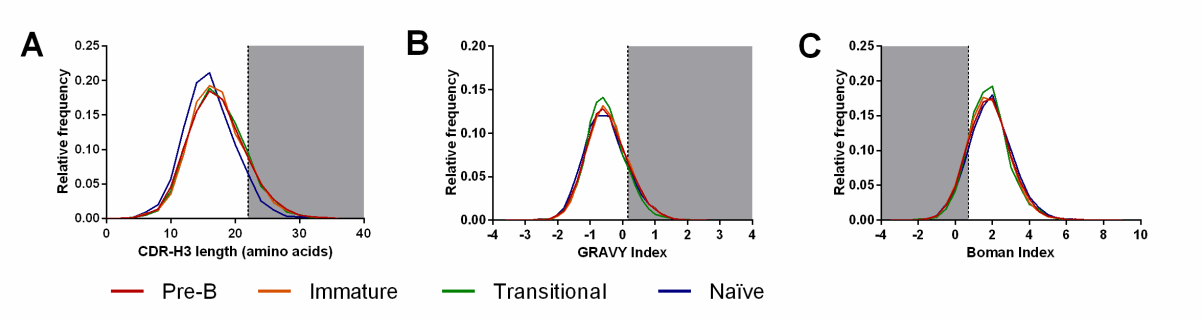


Figure 3-4: CDR-H3 property range of *Forbidden* antibodies. Histograms showing the values from which sequences were chosen for the (**A**) long (2 SD higher than mean), (**B**) high GRAVY (1 SD higher than mean) and (**C**) low Boman (1 SD lower than mean) CDR-H3. Graphs show the entire repertoire for each of the four cell types for all donors. The cut-offs were calculated using the naïve B cell repertoire of only the young donors (<65 years old).

When the B cell Development heavy chain dataset was filtered using these criteria, 62 heavy chain sequences remained (35 pre-B, 19 immature, four transitional and four naïve). 12 heavy chain sequences were selected (Table 3-2) from this group of 62 based on the fact that they used IGHV3 and IGHD2 genes that were less frequent in either the naïve or both the transitional and naïve repertoires compared to the pre-B and immature repertoires (IGHV3-15, IGHV3-30, IGHV3-33 and IGHD2-2) (Figure 3-1D&E).

As the selection of these heavy chains was due to their use of "forbidden" properties, this groups of heavy chains was named the *Forbidden* heavy chains.

Table 3-2: The 12 *Forbidden* IgH variable regions chosen for cloning.

IgH	Sequence ID	Donor	Cell Type	IGHV Gene	IGHD Gene	IGHJ Family	CDR-H3 amino acid sequence
1	immh103_M_UN_29876	103	Immature	IGHV3-15	IGHD2-2	IGHJ6	TTAPINIVVVPAAIKDRDYYYYGMDV
2	preh159_M_V3_60711	159	Pre-B	IGHV3-15	IGHD2-2	IGHJ6	ILVHIVVVPAAINGYYDFWSGWDV
3	naih111_M_V3_30355	111	Naïve	IGHV3-30	IGHD2-2	IGHJ6	AKTLLDVVVVPAATPQNYYYYGMDV
4	preh119_M_V3_110284	119	Pre-B	IGHV3-15	IGHD2-2	IGHJ6	TTHIVVVPAAIPSHYYYYYGMDV
5	preh149_M_UN_598450	149	Pre-B	IGHV3-30	IGHD2-2	IGHJ6	AKGAPNIVVVPAAKSAHYYYGMDV
6	preh118_M_V3_855230	118	Pre-B	IGHV3-33	IGHD2-2	IGHJ6	ARDTGVGGLVVPAATGDLAGYYYGMDV
7	immh103_M_V3_22024	103	Immature	IGHV3-15	IGHD2-2	IGHJ6	TTVRDIVVVPAAIGLNYYYGMDV
8	preh146_M_V3_253300	146	Pre-B	IGHV3-30	IGHD2-2	IGHJ6	AKTASFVVVPAAISDGDGYYYYMDV
9	immh149_M_V3_59726	149	Immature	IGHV3-30	IGHD2-2	IGHJ6	AKDQFVAVVVPAAIHLGYYYGMDV
10	preh149_M_UN_713160	149	Pre-B	IGHV3-15	IGHD2-2	IGHJ6	TTDRWGIVVVPAAKLVSYYGMDV
11	immh159_M_UN_59280	159	Immature	IGHV3-15	IGHD2-2	IGHJ6	TTPPGHIVVVPAANSLIAAPYYYYGMDV
12	preh138_M_V3_494120	138	Pre-B	IGHV3-15	IGHD2-2	IGHJ6	TTDLFLRARVVPAAMAPYYYGMDV

Each of the heavy chains in Table 3-2 was paired with three light chains representing a broad range of physicochemical properties. All three light chains were public light chains, identified in DeKosky et al. (2015) and they all appeared in the B cell Development dataset. Public light chains occur in multiple individuals, paired with various heavy chains. Two kappa and one lambda light chain were chosen and the three light chains had a range of CDR-L3 physicochemical properties, particularly the frequency of aromatic amino acids and aspartic acids in the CDR-L3 region as these features have been associated with rescuing autoreactive heavy chains (Li et al., 2001). The three light chains used are shown in Table 3-3.

Once all of the 12 *Forbidden* heavy chains had been paired with each of the three public light chains, a total of 36 *Forbidden* antibodies were expressed (see Table 3-4 on the subsequent two pages for details).

Table 3-3: Three public light chain variable region sequences. These light chains were chosen from the B cell Development dataset, however all their CDR-L3 nucleotide sequences also appear in a published list of public light chains (DeKosky et al., 2015).

IgL	Sequence ID	Donor	Cell Type	Isotype	IGLV Gene	IGLJ Family	CDR-L3 amino acid sequence
1	nail107_IGKV3_IGKC_111176	107	Naive	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
2	nail120_IGLC_UN_122004	120	Naive	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
3	imml103_IGKC_IGKV3_101790	103	Immature	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT

Table 3-4: 36 *Forbidden* antibodies properties. Name of the antibody followed by the B cell Development dataset sequence ID of the heavy and light chains and sequence information for each chain (table continues on next page).

Antibody ID	Heavy Chain ID	IGHV Gene	IGHD Gene	IGHJ Family	CDR-H3 amino acid sequence	Light Chain ID	Kappa/ Lambda	IGLV Gene	IGLJ Gene	CDR-L3 amino acid sequence
CT0001	imnh103_M_UN_29876	IGHV3-15	IGHD2-2	IGHJ6	TTAPINIVVVPAAIKDRDYYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0002	imnh103_M_UN_29876	IGHV3-15	IGHD2-2	IGHJ6	TTAPINIVVVPAAIKDRDYYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0003	imnh103_M_UN_29876	IGHV3-15	IGHD2-2	IGHJ6	TTAPINIVVVPAAIKDRDYYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0004	preh159_M_V3_60711	IGHV3-15	IGHD2-2	IGHJ6	ILVHIVVVPAAINGYYDFWSGWDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0005	preh159_M_V3_60711	IGHV3-15	IGHD2-2	IGHJ6	ILVHIVVVPAAINGYYDFWSGWDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0006	preh159_M_V3_60711	IGHV3-15	IGHD2-2	IGHJ6	ILVHIVVVPAAINGYYDFWSGWDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0007	naih111_M_V3_30355	IGHV3-30	IGHD2-2	IGHJ6	AKTLLDVVVVPAATPQNYYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0008	naih111_M_V3_30355	IGHV3-30	IGHD2-2	IGHJ6	AKTLLDVVVVPAATPQNYYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0009	naih111_M_V3_30355	IGHV3-30	IGHD2-2	IGHJ6	AKTLLDVVVVPAATPQNYYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0010	preh119_M_V3_110284	IGHV3-15	IGHD2-2	IGHJ6	TTHIVVVPAAIPSHYYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0011	preh119_M_V3_110284	IGHV3-15	IGHD2-2	IGHJ6	TTHIVVVPAAIPSHYYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0012	preh119_M_V3_110284	IGHV3-15	IGHD2-2	IGHJ6	TTHIVVVPAAIPSHYYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0013	preh149_M_UN_598450	IGHV3-30	IGHD2-2	IGHJ6	AKGAPNIVVVPAAKSAHYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0014	preh149_M_UN_598450	IGHV3-30	IGHD2-2	IGHJ6	AKGAPNIVVVPAAKSAHYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0015	preh149_M_UN_598450	IGHV3-30	IGHD2-2	IGHJ6	AKGAPNIVVVPAAKSAHYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0016	preh118_M_V3_855230	IGHV3-33	IGHD2-2	IGHJ6	ARDTGVGGLVVPAATGDLAGYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0017	preh118_M_V3_855230	IGHV3-33	IGHD2-2	IGHJ6	ARDTGVGGLVVPAATGDLAGYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0018	preh118_M_V3_855230	IGHV3-33	IGHD2-2	IGHJ6	ARDTGVGGLVVPAATGDLAGYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT

Antibody ID	Heavy Chain ID	IGHV Gene	IGHD Gene	IGHJ Family	CDR-H3 amino acid sequence	Light Chain ID	Kappa/ Lambda	IGLV Gene	IGLJ Gene	CDR-L3 amino acid sequence
CT0019	immh103_M_V3_22024	IGHV3-15	IGHD2-2	IGHJ6	TTVRDIVVVPAAIGLNYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0020	immh103_M_V3_22024	IGHV3-15	IGHD2-2	IGHJ6	TTVRDIVVVPAAIGLNYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0021	immh103_M_V3_22024	IGHV3-15	IGHD2-2	IGHJ6	TTVRDIVVVPAAIGLNYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0022	preh146_M_V3_253300	IGHV3-30	IGHD2-2	IGHJ6	AKTASFVVVPAAISDGDGYYYYMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0023	preh146_M_V3_253300	IGHV3-30	IGHD2-2	IGHJ6	AKTASFVVVPAAISDGDGYYYYMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0024	preh146_M_V3_253300	IGHV3-30	IGHD2-2	IGHJ6	AKTASFVVVPAAISDGDGYYYYMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0025	immh149_M_V3_59726	IGHV3-30	IGHD2-2	IGHJ6	AKDQFVAVVVPAAIHLGYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0026	immh149_M_V3_59726	IGHV3-30	IGHD2-2	IGHJ6	AKDQFVAVVVPAAIHLGYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0027	immh149_M_V3_59726	IGHV3-30	IGHD2-2	IGHJ6	AKDQFVAVVVPAAIHLGYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0028	preh149_M_UN_713160	IGHV3-15	IGHD2-2	IGHJ6	TTDRWGIVVVPAAKLVSYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0029	preh149_M_UN_713160	IGHV3-15	IGHD2-2	IGHJ6	TTDRWGIVVVPAAKLVSYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0030	preh149_M_UN_713160	IGHV3-15	IGHD2-2	IGHJ6	TTDRWGIVVVPAAKLVSYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0031	immh159_M_UN_59280	IGHV3-15	IGHD2-2	IGHJ6	TTPPGHIVVVPAANSLIAAPYYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0032	immh159_M_UN_59280	IGHV3-15	IGHD2-2	IGHJ6	TTPPGHIVVVPAANSLIAAPYYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0033	immh159_M_UN_59280	IGHV3-15	IGHD2-2	IGHJ6	TTPPGHIVVVPAANSLIAAPYYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT
CT0034	preh138_M_V3_494120	IGHV3-15	IGHD2-2	IGHJ6	TTDLFLRARVVPAAMAPYYYGMDV	nail107_IGKV3_IGKC_111176	Kappa	IGKV3-20	IGKJ2	QQYGSSPYT
CT0035	preh138_M_V3_494120	IGHV3-15	IGHD2-2	IGHJ6	TTDLFLRARVVPAAMAPYYYGMDV	nail120_IGLC_UN_122004	Lambda	IGLV1-44	IGLJ2	AAWDDSLNGVV
CT0036	preh138_M_V3_494120	IGHV3-15	IGHD2-2	IGHJ6	TTDLFLRARVVPAAMAPYYYGMDV	imml103_IGKC_IGKV3_101790	Kappa	IGKV3-11	IGKJ4	QQRSNWPLT

3.3.2 Selection of *Long Hydrophobic* Heavy Chains

The properties used to select the *Forbidden* heavy chains (section 3.3.1) were quite extreme and the resulting heavy chains were quite similar. Therefore, the heavy chains selected in this section had fewer constraints. The B cell Development dataset was filtered for entries which met both of these criteria:

1. CDR-H3 >15.5 amino acids long
2. CDR-H3 GRAVY index >-0.521

The cut-offs for these properties were defined as being higher than the mean CDR-H3 length and mean GRAVY index in the young naïve population. The sequence with the highest GRAVY index and the sequence with the longest CDR-H3 were selected. The other three sequences were chosen semi-randomly (Table 3-5).

All three *Long Hydrophobic* heavy chains were paired with the same light chain (see Table 3-5).

Table 3-5: Properties of five *Long Hydrophobic* heavy chains and the light chain with which they were all paired.

Antibody ID	Heavy chain ID	Donor	Cell Type	IGHV gene	IGHJ gene	IGHD gene	CDR-H3 AA sequence	GRAVY Index
CT0040	immh141_M_V4_51077	141	Immature	IGHV4-39	IGHJ3	IGHD3-3	ARELGFGVVIILVLLIS	2.135
CT0041	immh160_M_UN_110852	160	Immature	IGHV4-61	IGHJ2	IGHD4-17	ASLYGDYVPTDWYFDL	-0.150
CT0050	preh111_M_V3_567190	111	Pre-B	IGHV3-30	IGHJ3	IGHD2-15	AKVGGQDIVVVVAALDI	1.412
CT0053	preh138_M_UN_483280	138	Pre-B	IGHV3-23	IGHJ4	IGHD3-16	AKDIVVLHLTAADTAVYYCARAGGGLRLGELSLLY	0.649
CT0054	preh146_M_V3_113150	146	Pre-B	IGHV3-15	IGHJ3	IGHD3-10	TTDPSNYGITMVQGVIITSSNRDAFDI	-0.056
	Light chain ID	Donor	Cell Type	IGKV gene	IGKJ gene		CDR-L3 AA sequence	
	nail119_IGKV3_UN_122382	119	Naïve	IGKV1-37	IGKJ5		QQYNNWPRT	

3.3.3 Selection of *Preferred* Heavy Chains

The 12 *Forbidden* heavy chains were chosen because they had properties which decreased in frequency as B cells matured as it was hypothesised that they would be more likely to be autoreactive/polyreactive. Therefore, for comparison, five *Preferred* heavy chains were selected which had properties which increased in frequency as the B cells matured. It was hypothesised that these heavy chains should not show autoreactivity or polyreactivity as they possess properties which are not removed from the repertoire as B cells mature.

The B cell Development dataset was filtered for entries which met all of the following criteria:

1. CDR-H3 <15.5 amino acids long
2. CDR-H3 GRAVY index <-0.522
3. CDR-H3 small amino acid frequency <0.559
4. CDR-H3 Boman index >1.93
5. CDR-H3 aromatic amino acid frequency >0.215
6. CDR-H3 charged amino acid frequency >0.251
7. CDR-H3 acidic amino acid >0.136

The cut-off values chosen for each of these properties was the mean value in the young naïve population. Five sequences were chosen semi-randomly, ensuring that a variety of VDJ combinations were represented (see Table 3-6).

Table 3-6: Properties of the five *Preferred* heavy chains selected and the light chain with which they were all paired.

Antibody ID	Heavy chain ID	Donor	Cell Type	IGHV Gene	IGHJ Gene	IGHD Gene	CDR-H3 AA sequence	GRAVY Index	Boman Index
CT0042	immh160_M_V5_26643	160	Immature	IGHV5-51	IGHJ4	IGHD4-17	ARLSGDYRFDY	-1.036	3.665
CT0044	naih141_M_V4_13619	141	Naïve	IGHV4-59	IGHJ4	IGHD7-27	ARDHQLGMFDY	-0.873	2.700
CT0052	preh119_M_V3_526840	119	Pre-B	IGHV3-23	IGHJ1	IGHD3-16	AREWGSYGNEYFQH	-1.586	2.861
CT0058	trah119_M_V4_55123	119	Transitional	IGHV4-31	IGHJ3	IGHD3-9	ARDHFDWDDAFDI	-0.954	3.565
CT0059	trah141_M_V3_25243	141	Transitional	IGHV3-23	IGHJ6	IGHD1-1	AKPREDYYYYYYMDV	-1.360	2.491
	Light chain ID	Donor	Cell Type	IGKV gene	IGKJ gene		CDR-L3 AA sequence		
	nail119_IGKV3_UN_122382	119	Naïve	IGKV1-37	IGKJ5		QQYNNWPRT		

3.3.4 Selection of *Aromatic* Heavy Chains

There is some evidence from the literature (Droupadi, 1994) and work by other members of the Dunn-Walters lab that aromatic amino acid content of the CD-H3 region may be related to polyreactivity. Therefore, the B cell Development dataset was filtered for entries which met the following criteria:

1. CDR-H3 aromatic amino acid frequency >0.315
2. All other measured CDR-H3 physicochemical properties within 1 SD of the mean

The cut-off was 1 SD greater than the mean aromatic amino acid frequency in the young naïve population. The five heavy chains selected had aromatic amino acids spread across the CDR-H3 and it was ensured that a variety of VDJ genes were represented (Table 3-7).

Table 3-7: Properties of the five *Aromatic* heavy chains selected which have a high frequency of aromatic amino acids in the CDR-H3. Aromatic amino acids (F/H/W/Y) are underlined. Also the light chain with which they were all paired.

Antibody ID	Heavy chain ID	Donor	Cell Type	IGHV Gene	IGHJ Gene	IGHD Gene	CDR-H3 AA sequence
CT0037	immh119_M_V3_21823	119	Immature	IGHV3-30	IGHJ4	IGHD5-12	AGHDSGYDWPYFLY
CT0045	naih140_M_V3_60340	140	Naïve	IGHV3-48	IGHJ4	IGHD3-22	ARVGYYYDSSGYYFDY
CT0047	naih160_M_V3_30414	160	Naïve	IGHV3-23	IGHJ2	IGHD3-10	AKDAYGSGSYYNWYFDL
CT0048	naih160_M_V4_11003	160	Naïve	IGHV4-61	IGHJ2	IGHD6-19	ARDGSGWYGDWYFDL
CT0060	trah146_UN_V3_28720	146	Transitional	IGHV3-21	IGHJ6	IGHD6-6	ARDHAYSSSMDYWYYGMDV
	Light chain ID	Donor	Cell Type	IGKV gene	IGKJ gene		CDR-L3 AA sequence
	nail119_IGKV3_UN_122382	119	Naïve	IGKV1-37	IGKJ5		QQYNNWPRT

3.3.5 Selection of *pI 5.0-6.0* Heavy Chains

It has been speculated that the frequency of charged amino acids and isoelectric points (pI) could be linked to polyreactivity (Zhang & Yeh, 1994). The histograms shown in Figure 3-2D (page 142) show that there is a very low frequency of CDR-H3 with a pI in the range 5.0-6.0 in all four cell types. We speculated that the reason for this may be that CDR-H3 with a pI in this region are more prone to autoreactivity/polyspecificity. The B cell Development dataset was filtered for entries which met the following criterion:

1. CDR-H3 pI between 5.0 and 6.0

Of the entries with met this criterion, the entry with highest pI (5.996) and the lowest pI (5.250) were selected. The other three were chosen semi-randomly, making sure that a variety of VDJ genes were represented (Table 3-8).

Table 3-8: Properties of five *pI 5.0-6.0* heavy chains selected which have a CDR-H3 with an isoelectric point within the range 5.0 to 6.0 and the light chain with which they were all paired.

Antibody ID	Heavy chain ID	Donor	Cell Type	IGHV Gene	IGHJ Gene	IGHD Gene	CDR-H3 AA sequence	pI
CT0039	immh141_M_V2_123166	141	Immature	IGHV2-5	IGHJ4	IGHD3-16	AHSYYDYVWGSYRTYDP	5.408
CT0043	naih119_M_V6_23731	119	Naïve	IGHV6-1	IGHJ4	IGHD3-16	ARGVHDYVWGSPLGFDY	5.408
CT0049	preh103_M_V3_555440	103	Pre-B	IGHV3-23	IGHJ5	IGHD2-15	AKGCNCSGGSCYSGWFDP	5.996
CT0057	preh149_M_V2_368690	149	Pre-B	IGHV2-5	IGHJ5	IGHD3-9	AHLHILTGLDCGFDP	5.250
CT0061	trah159_M_V3_33486	159	Transitional	IGHV3-74	IGHJ6	IGHD2-15	ARDCTGAIGCHYGMDV	5.403
	Light chain ID	Donor	Cell Type	IGKV gene	IGKJ gene		CDR-L3 AA sequence	
	nail119_IGKV3_UN_122382	119	Naïve	IGKV1-37	IGKJ5		QQYNNWPRT	

3.3.6 Selection of *pI 7.0-8.5* Heavy Chains

The rationale for these heavy chains is the same as in section 3.3.5. There is a very low frequency of heavy chain sequences with a CDR-H3 pI of between 7.0-8.5 in all four cell types (Figure 3-2D). The B cell Development dataset was filtered for entries which met the following criterion:

1. CDR-H3 pI between 7.0 and 8.5

Of the entries which met this criterion, the entries with the highest pI (8.449) and the lowest pI (7.173) were chosen. The other three were chosen semi-randomly, making sure that a variety of VDJ genes were represented (Table 3-9).

Table 3-9: Properties of five *pI 7.0-8.5* heavy chains selected which have a CDR-H3 with an isoelectric point within the range 7.0 to 8.5 and the light chain with which they were all paired.

Antibody ID	Heavy chain ID	Donor	Cell Type	IGHV Gene	IGHJ Gene	IGHD Gene	CDR-H3 AA sequence	pI
CT0038	immh119_M_V3_64902	119	Immature	IGHV3-11	IGHJ4	IGHD2-2	ARPYCSSTSCPANY	8.215
CT0046	naih140_M_V3_70569	140	Naïve	IGHV3-11	IGHJ6	IGHD2-2	AKCMGNCSSTSCYTFGHYYYGMDV	7.173
CT0051	preh119_M_V3_279470	119	Pre-B	IGHV3-30	IGHJ4	IGHD2-2	AKDYWLRGKPDCSSTSPRSSWFDY	8.499
CT0055	preh146_M_V3_220600	146	Pre-B	IGHV3-7	IGHJ6	IGHD5-24	ARVPPTHYYMDV	7.535
CT0056	preh146_M_V3_592310	146	Pre-B	IGHV3-23	IGHJ3	IGHD3-3	ARVHSSRDFWSGYYVGYGAFDI	7.528
	Light chain ID	Donor	Cell Type	IGKV gene	IGKJ gene		CDR-L3 AA sequence	
	nail119_IGKV3_UN_122382	119	Naïve	IGKV1-37	IGKJ5		QQYNNWPRT	

3.4 Investigating Autoreactivity against HEp-2 Cells

To determine if any of the antibodies selected in section 3.3 (Identification of Potentially Autoreactive Heavy Chains) are autoreactive, all of the antibodies which could be expressed were tested for binding to HEp-2 cells using immunocytochemistry. HEp-2 cells are commonly used in the clinical diagnosis of autoimmune disease (Sack et al., 2009, Buchner et al., 2014). Briefly, HEp-2 cells were cultured on glass coverslips, fixed with paraformaldehyde and blocked with BSA. The fixed cells were incubated with purified primary IgG1 antibody (section 3.3) and then fluorescently-labelled anti-human IgG secondary antibody. Slides were imaged using 3 µm Z-stacks on a widefield fluorescence microscope. DAPI was used to confirm the presence of cells in every image. For full materials and methods see section 2.9 (page 109).

3.4.1 *Forbidden* Antibodies

Two of the *Forbidden* antibodies (CT0035 and CT0036) could not be expressed. IgH chain 12 could only be expressed when in combination with one of the kappa light chains.

Of the 34 *Forbidden* antibodies which could be expressed, 11 were strong HEp-2 binders, 13 were weak HEp-2 binders and the remaining 10 did not bind HEp-2 cells. For results, see Figure 3-5 on the following three pages and Table 3-10 on page 157.

Of the HEp-2 binders, five bound both the nucleus and the cytoplasm, and the remaining 19 bound the cytoplasm only (none bound the nucleus only). Six heavy chains (IgH chains 3, 4, 5, 6, 9 and 11; Table 3-10, page 160) showed the same binding patterns with all three public light chains. IgH chains 1 and 7 both bound the HEp-2 cells when paired with the kappa light chains, but not when paired with the lambda light chain. The binding patterns of antibodies with IgH chains 2, 8 and 10 also differed depending on the light chain with which they were paired.

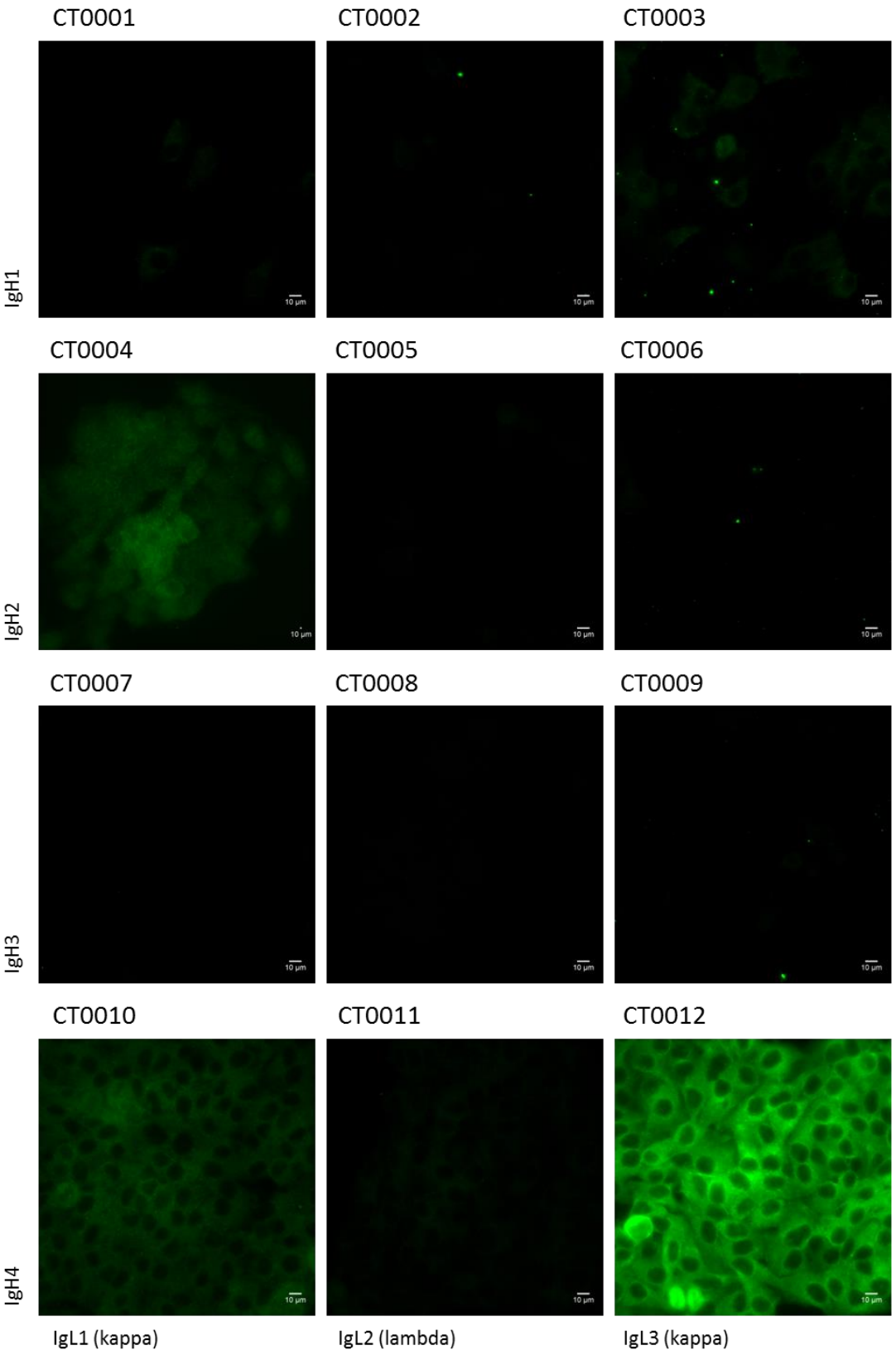


Figure 3-5 continues on next page. See page 156 for caption.

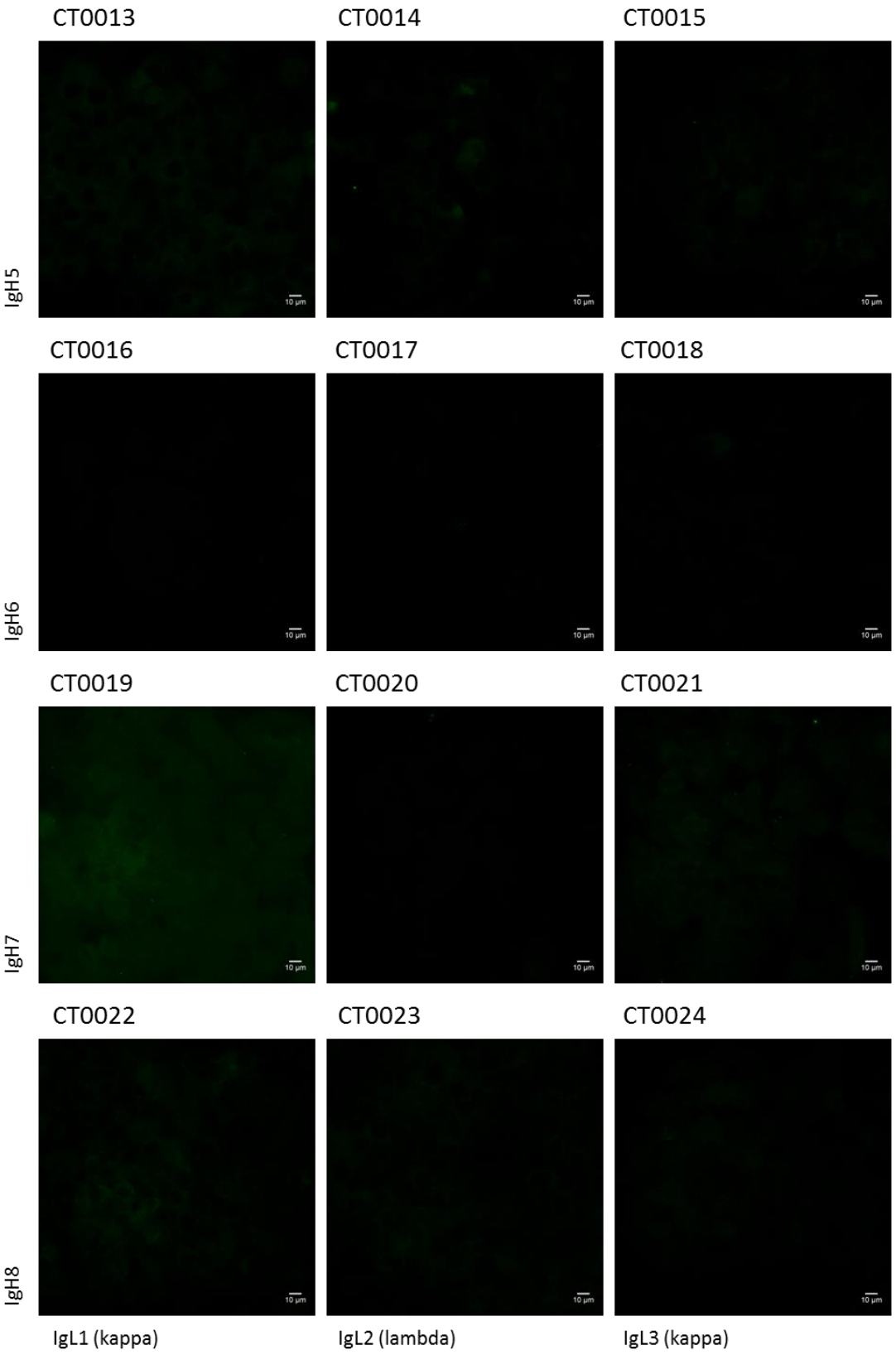


Figure 3-5 continues on next page. See page 156 for caption.

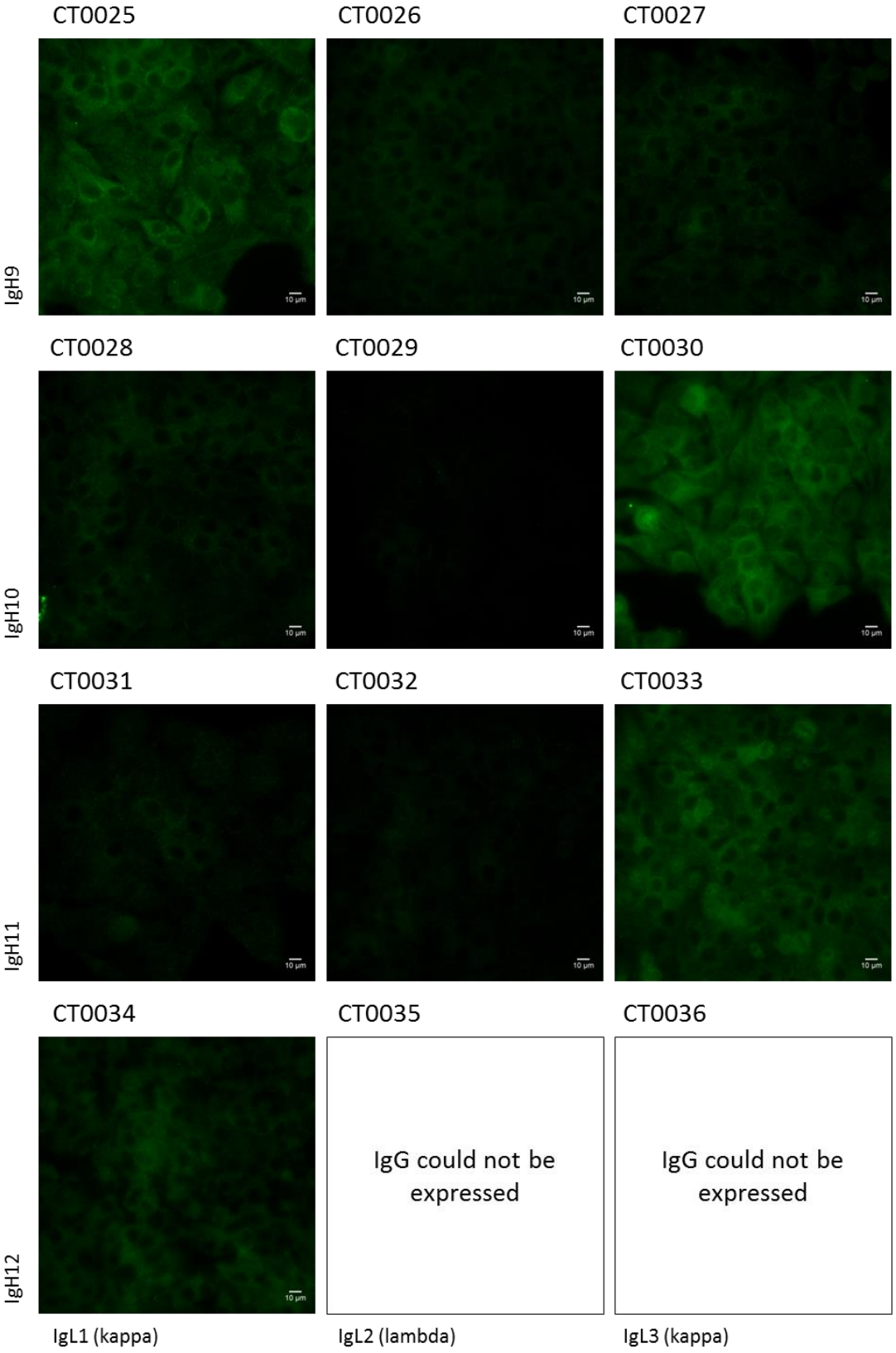


Figure 3-5: HEp-2 immunocytochemistry for the 36 *Forbidden* antibodies. All images were taken using 0.3µm Z-stacks and the exposure was 100 ms. Images were analysed in ImageJ. DAPI was used to ensure cells were present in every image. Heavy and light chain numbers are as in Table 3-2 (page 145) and Table 3-3 (page 146).

Table 3-10: HEp-2 binding of 36 *Forbidden* antibodies. Refers to results from Figure 3-5. Dotted lines group the antibodies with the same heavy chain. Each heavy chain has been given a number in the IgH column. The class of light chain is shown in the IgL column. Details of each antibody are given in Table 3-4 (page 147).

Antibody ID	IgH	IgL	HEp-2 binding	Cytoplasmic binding	Nuclear binding
CT0001		1 (κ)	Weak	✓	×
CT0002	1	2 (λ)	No binding	×	×
CT0003		3 (κ)	Weak	✓	×
CT0004		1 (κ)	Strong	✓	✓
CT0005	2	2 (λ)	No binding	×	×
CT0006		3 (κ)	No binding	×	×
CT0007		1 (κ)	No binding	×	×
CT0008	3	2 (λ)	No binding	×	×
CT0009		3 (κ)	No binding	×	×
CT0010		1 (κ)	Strong	✓	×
CT0011	4	2 (λ)	Weak	✓	×
CT0012		3 (κ)	Strong	✓	×
CT0013		1 (κ)	Weak	✓	×
CT0014	5	2 (λ)	Weak	✓	×
CT0015		3 (κ)	Weak	✓	×
CT0016		1 (κ)	No binding	×	×
CT0017	6	2 (λ)	No binding	×	×
CT0018		3 (κ)	No binding	×	×
CT0019		1 (κ)	Strong	✓	✓
CT0020	7	2 (λ)	No binding	×	×
CT0021		3 (κ)	Weak	✓	✓
CT0022		1 (κ)	Weak	✓	×
CT0023	8	2 (λ)	Weak	✓	×
CT0024		3 (κ)	Weak	✓	✓
CT0025		1 (κ)	Strong	✓	×
CT0026	9	2 (λ)	Strong	✓	×
CT0027		3 (κ)	Strong	✓	×
CT0028		1 (κ)	Strong	✓	×
CT0029	10	2 (λ)	Weak	✓	×
CT0030		3 (κ)	Strong	✓	✓
CT0031		1 (κ)	Weak	✓	×
CT0032	11	2 (λ)	Weak	✓	×
CT0033		3 (κ)	Strong	✓	×
CT0034		1 (κ)	Strong	✓	×
CT0035	12	2 (λ)	Not Expressed	NA	NA
CT0036		3 (κ)	Not Expressed	NA	NA

3.4.2 *Long Hydrophobic* Antibodies

The *Long Hydrophobic* antibodies all had heavy chains with above-average CDR-H3 length and hydrophobicity.

Of the five *Long Hydrophobic* antibodies, three bound HEp-2 cells. CT0050 showed weak nuclear binding and CT0053 and CT0054 showed strong cytoplasmic binding. CT0041 did not bind HEp-2 cells and CT0040 could not be expressed. See Figure 3-6 and Table 3-11 for results.

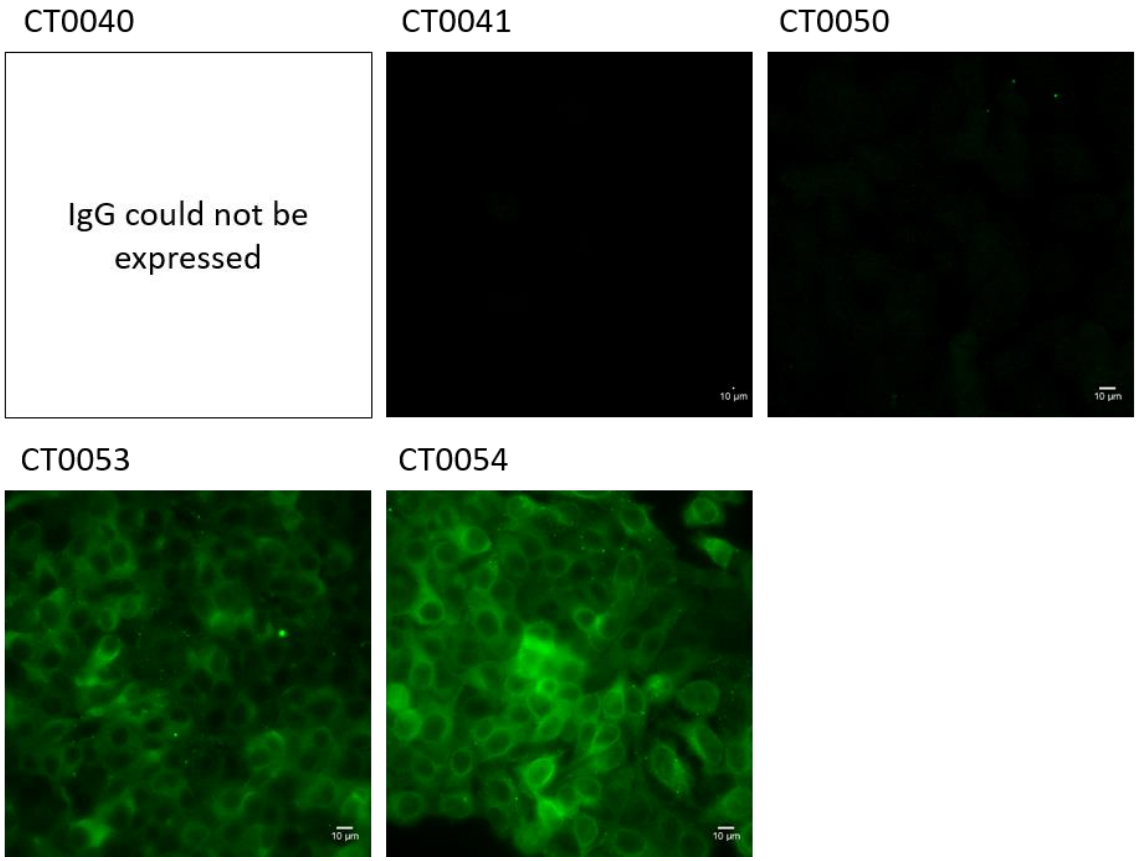


Figure 3-6: HEp-2 immunocytochemistry for the five *Long Hydrophobic* antibodies.

Table 3-11: HEp-2 binding of five *Long Hydrophobic* antibodies (antibodies with long, hydrophobic CDR-H3). Refers to results from Figure 3-6. All five antibodies have the same light chain. Details of each antibody are given in Table 3-5 (page 149).

Antibody ID	IgL	HEp-2 binding	Cytoplasmic binding	Nuclear binding
CT0040	κ	Not Expressed	NA	NA
CT0041	κ	No binding	✗	✗
CT0050	κ	Weak	✗	✓
CT0053	κ	Strong	✓	✗
CT0054	κ	Strong	✓	✗

3.4.3 *Preferred* Antibodies

The *Preferred* antibodies all had CDR-H3 properties which increased in prevalence as B cells matured, indicating that they may be less likely to be autoreactive.

Of the five *Preferred* antibodies, CT0042, CT0044 and CT0052 showed no binding to HEp-2 cells. CT0058 and CT0059 showed weak cytoplasmic and nuclear binding. All of the *Preferred* antibodies could be expressed. See Figure 3-7 and Table 3-12 for results.

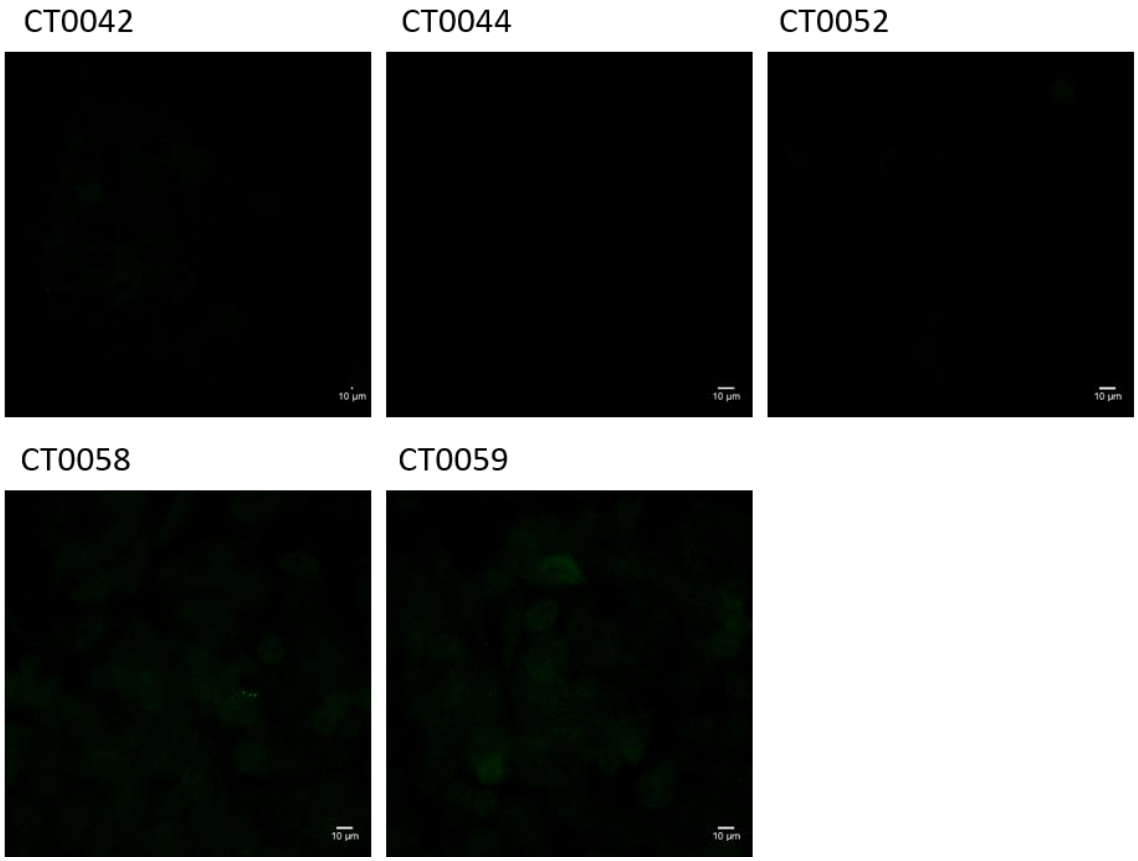


Figure 3-7: HEp-2 immunocytochemistry for the five *Preferred* antibodies.

Table 3-12: HEp-2 binding of five *Preferred* antibodies. Refers to results from Figure 3-7. All five antibodies have the same light chain. Details of the antibodies are given in Table 3-6 (page150).

Antibody ID	IgL	HEp-2 binding	Cytoplasmic binding	Nuclear binding
CT0042	κ	No binding	×	×
CT0044	κ	No binding	×	×
CT0052	κ	No binding	×	×
CT0058	κ	Weak	✓	✓
CT0059	κ	Weak	✓	✓

3.4.4 *Aromatic* Antibodies

The *Aromatic* antibodies all had very high frequencies (>1 SD above the mean) of aromatic amino acids in the CDR-H3.

CT0037 showed very strong cytoplasmic binding, but no evidence of nuclear binding. CT0047 and CT0060 showed weak cytoplasmic binding. CT0045 did not bind HEp-2 cells and CT0048 could not be expressed. See Figure 3-8 and Table 3-13 for results.

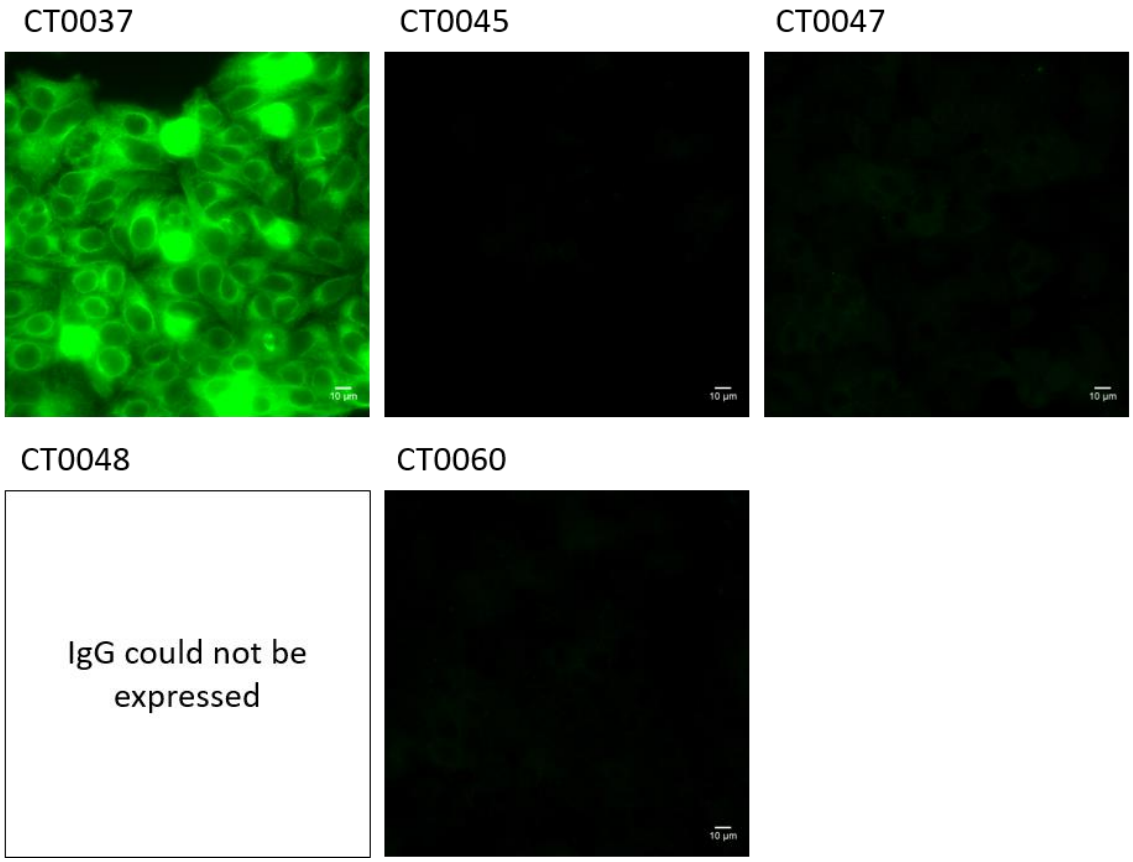


Figure 3-8: HEp-2 immunocytochemistry for the five *Aromatic* antibodies.

Table 3-13: HEp-2 binding of five *Aromatic* antibodies (CDR-H3 contains high proportion of aromatic amino acids). Refers to results from Figure 3-8. All five antibodies have the same light chain. Details of the antibodies are given in Table 3-7 (page 151).

Antibody ID	IgL	HEp-2 binding	Cytoplasmic binding	Nuclear binding
CT0037	κ	Strong	✓	×
CT0045	κ	No binding	×	×
CT0047	κ	Weak	✓	×
CT0048	κ	Not Expressed	NA	NA
CT0060	κ	Weak	✓	×

3.4.5 *pI 5.0-6.0* Antibodies

The *pI 5.0-6.0* antibodies all had CDR-H3 isoelectric points (pI) between 5.0 and 6.0. This pI range is very underused in all cell types measured.

CT0039 and CT0043 bound only the nucleus and CT0049 bound both the nucleus and the cytoplasm. CT0057 did not bind HEp-2 cells and CT0061 could not be expressed. See Figure 3-9 and Table 3-14 for results.

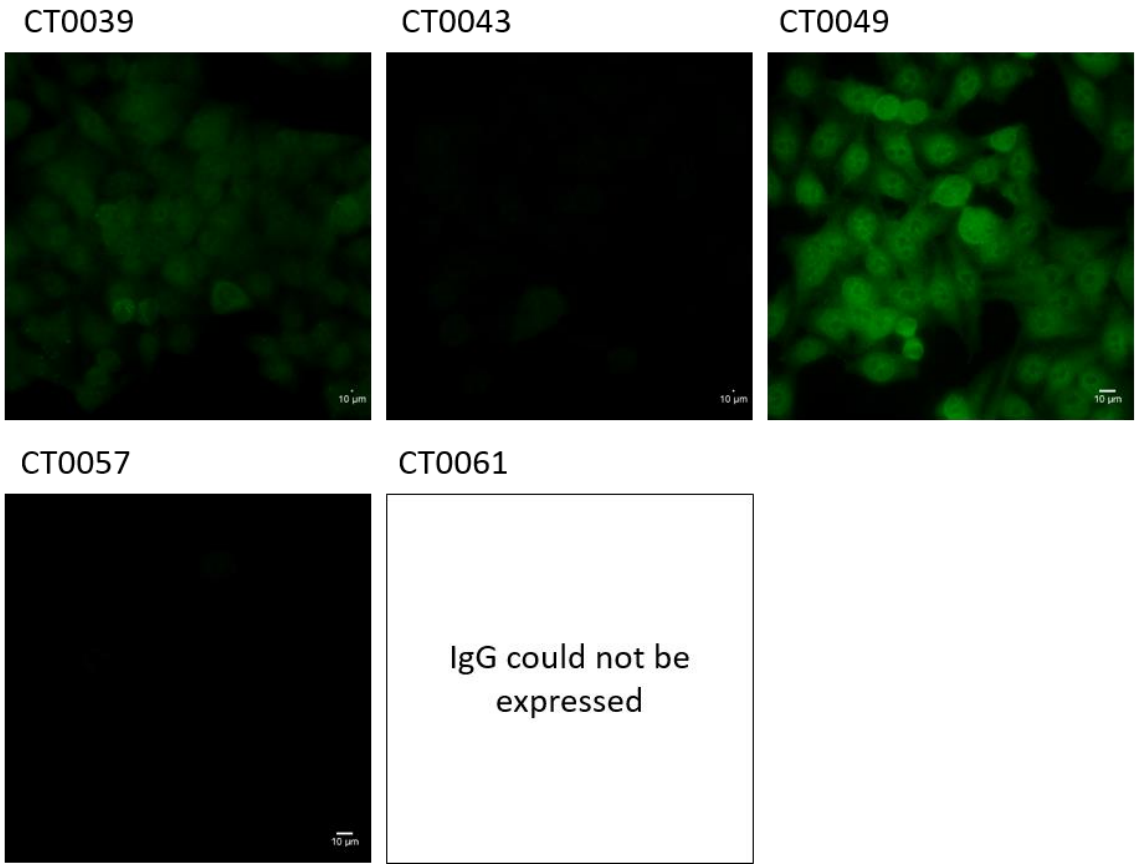


Figure 3-9: HEp-2 immunocytochemistry for the five *pI 5-6* antibodies.

Table 3-14: HEp-2 binding of five *pI 5.0-6.0* antibodies (CDR-H3 has a pI of between 5.0 and 6.0). Refers to results from Figure 3-9. All five antibodies have the same light chain. Details of the antibodies are given in Table 3-8 (page 151).

Antibody ID	IgL	HEp-2 binding	Cytoplasmic binding	Nuclear binding
CT0039	κ	Strong	✗	✓
CT0043	κ	Weak	✗	✓
CT0049	κ	Strong	✓	✓
CT0057	κ	No binding	✗	✗
CT0061	κ	Not Expressed	NA	NA

3.4.6 *pI 7.0-8.5* Antibodies

The *pI 7.0-8.5* antibodies all had CDR-H3 isoelectric points (pI) between 7.0 and 8.5. This pI range is very underused in all cell types measured.

CT0046, CT0051 and CT0056 strongly bind both the nucleus and the cytoplasm. CT0055 weakly binds the cytoplasm. CT0038 does not bind HEp-2 cells. All of the *pI 7.0-8.5* antibodies could be expressed. See Figure 3-10 and Table 3-15 for results.

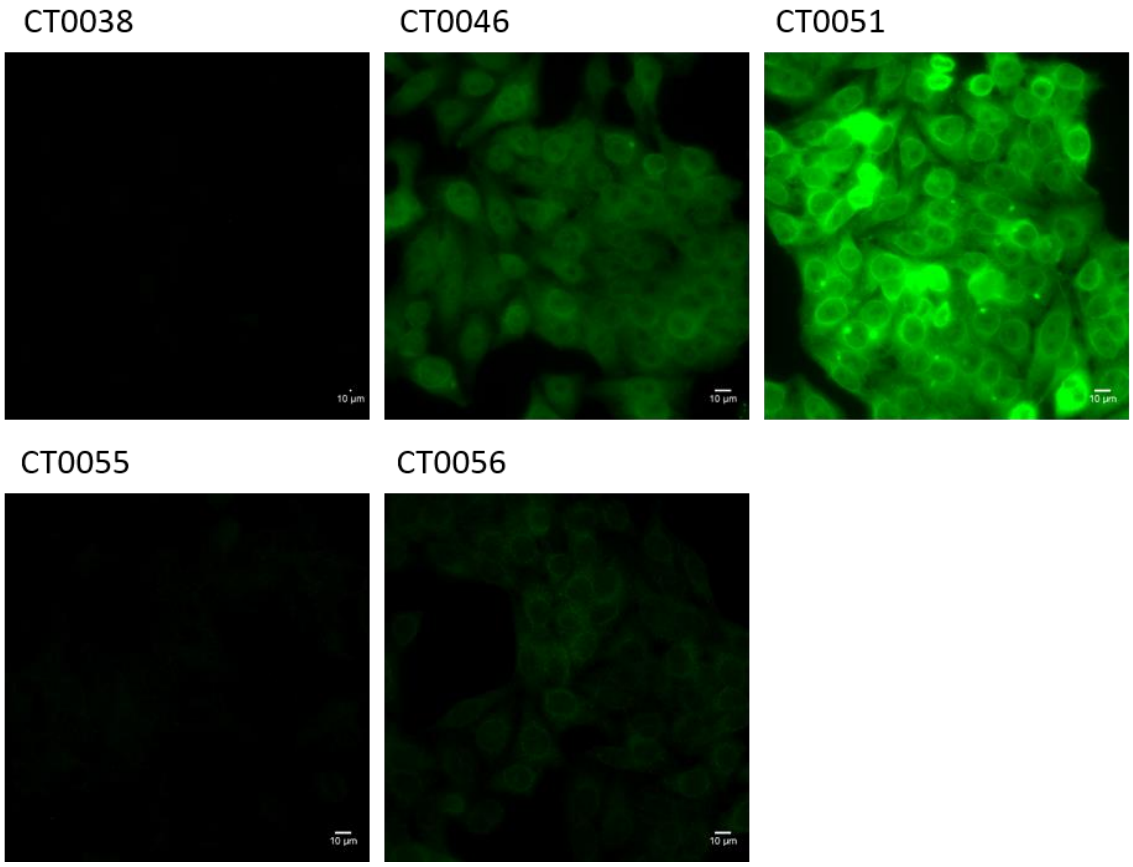


Figure 3-10: HEp-2 immunocytochemistry for the five *pI 7-8.5* antibodies.

Table 3-15: HEp-2 binding of five *pI 7.0-8.5* antibodies (CDR-H3 has a pI of between 7.0 and 8.5). Refers to results from Figure 3-10. All five antibodies have the same light chain. Details of the antibodies are given in Table 3-9 (page 152).

Antibody ID	IgL	HEp-2 binding	Cytoplasmic binding	Nuclear binding
CT0038	κ	No binding	✗	✗
CT0046	κ	Strong	✓	✓
CT0051	κ	Strong	✓	✓
CT0055	κ	Weak	✓	✗
CT0056	κ	Strong	✓	✓

3.5 Investigating Polyreactivity against a Panel of Unrelated Antigens

It is thought that autoreactivity and polyreactivity are closely linked; if an antibody is polyreactive and binds multiple epitopes, it is likely that it will bind self-antigen. One method of investigating polyreactivity is to test the antibodies in an ELISA for binding against multiple antigens (Wardemann et al., 2003, Tiller et al., 2008). If an antibody binds to more than one antigen, it is considered polyreactive.

3.5.1 Control Polyreactive Antibodies from the Literature

The protocol to test antibodies for polyreactivity was set out in Wardemann et al. (2003) and Tiller et al. (2008). To confirm that the polyreactivity ELISA protocol worked in our lab, the following known polyreactive positive and negative controls were expressed[89,90]:

1. ED38 - high polyreactive (Meffre et al., 2004)
2. ei-JB40 - low polyreactive (Wardemann et al., 2003)
3. m-GO53 - negative (Wardemann et al., 2003)

The specificity of the three control antibodies was tested against dsDNA, insulin, LPS, HEp-2 cell lysate and also blank negative control wells coated with PBS only. The ELISA protocol set out in Tiller et al. (2008) was followed. In this protocol, wells were blocked with 0.05% (v/v) Tween-20 2 mM EDTA in PBS only (no protein block).

Binding patterns of ED38, ei-JB40 and m-GO53 reported in (Tiller et al., 2008) could be replicated for dsDNA, LPS, insulin and HEp-2 cell lysate. However, the same binding patterns were also observed for wells coated with no antigen (PBS only) (Figure 3-11).

[89] Plasmids encoding ED38, ei-JB40 and m-GO53 were generously shared by Hedda Wardemann (German Cancer Research Center, Heidelberg, Germany)

[90] ED38, ei-JB40 and m-GO53 antibodies were expressed using the method described in section 2.8 (page 103)

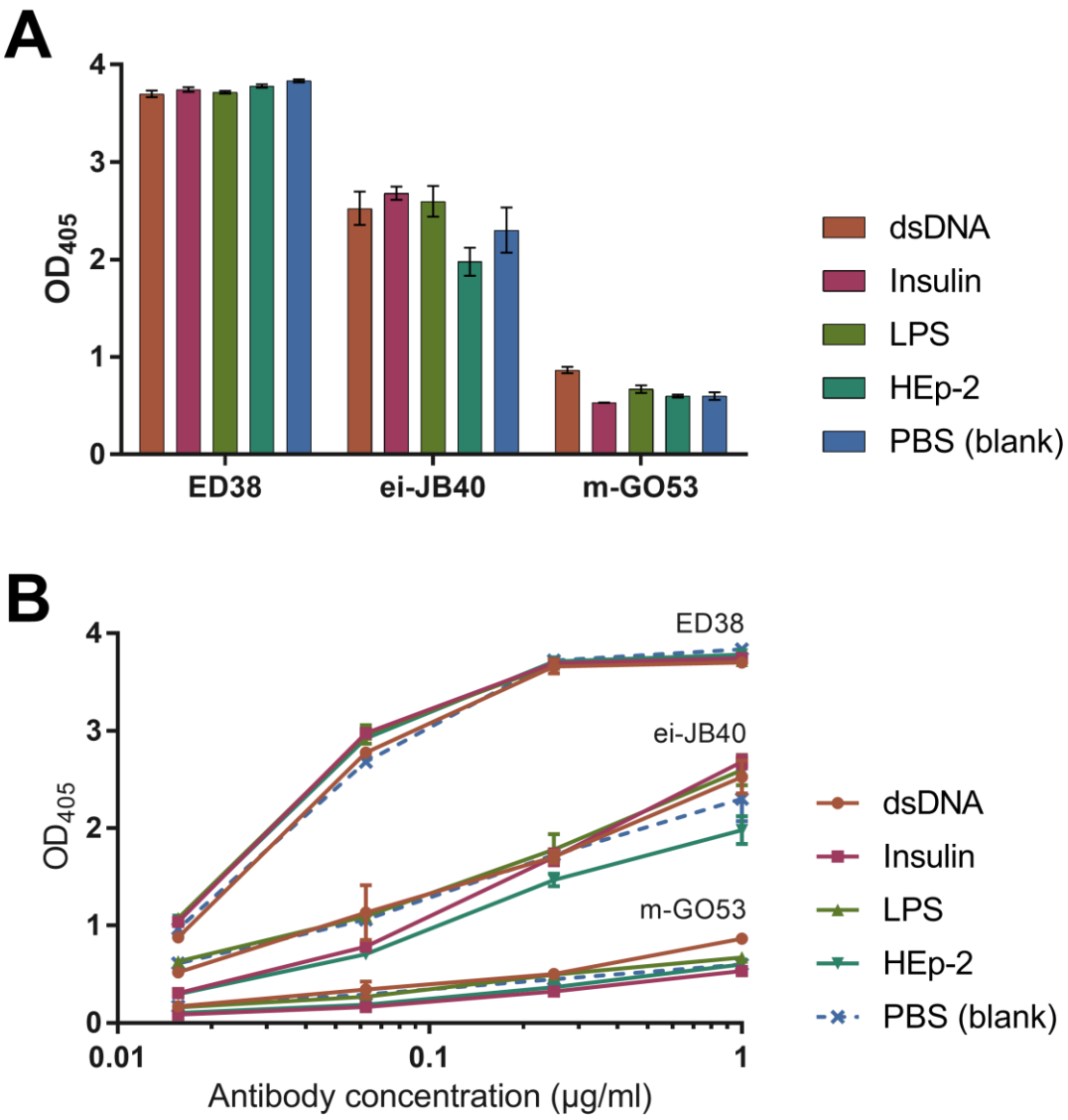


Figure 3-11: ELISA to test specificity of polyspecific control antibodies. All antibodies bind PBS-coated negative control wells just as well as they bind antigen-coated wells. (**A**) The OD_{405} when each antibody is at a concentration of 1µg/ml. (**B**) A replicate of the graphs used in the literature to show that the graphs follow the same curve in the absence of antibody as they do in the presence of antibody. Each graph shows the mean OD_{405} (+/- 1 SD) of each antibody tested in triplicate.

In light of the PBS (blank) result in Figure 3-11, we investigated whether the phenotype displayed by ED38 and ei-JB40 may be due to the unusual blocking conditions used in the Tiller et al. (2008) protocol, four blocking conditions were investigated:

1. 3% (w/v) milk powder in PBS (Figure 3-12A)
2. 3% (w/v) BSA in PBS (Figure 3-12B)
3. 0.05% (v/v) Tween-20 2 mM EDTA in PBS (Figure 3-12C)
4. 1% (v/v) Tween-20 0.3 M NaCl in 50 mM Tris buffer (pH 7.5) (Figure 3-12D)

To save time and resources only dsDNA and insulin were used as antigens alongside the blank (PBS only) wells. Anti-dsDNA and anti-insulin commercial positive controls were included to confirm that antigen-specific antibodies could bind under all four blocking conditions. Antigens and positive controls are detailed in Table 2-34 and Table 2-35 respectively (page 111).

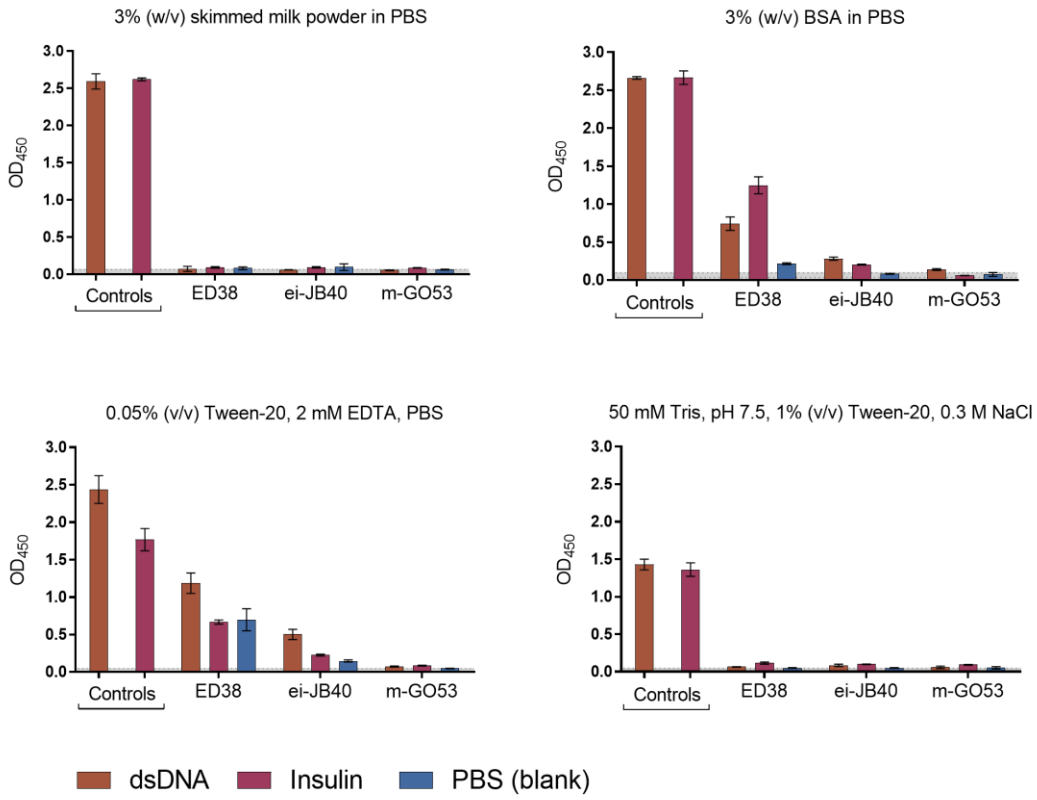


Figure 3-12: Four blocking conditions to test specificity of control antibodies. Blocking conditions used were: 3% (w/v) skimmed milk powder in PBS, 3% (w/v) BSA in PBS, 0.05% (v/v) Tween-20, 2 mM EDTA, PBS and 50 mM Tris, pH 7.5, 1% (v/v) Tween-20, 0.3 M NaCl. Mean OD_{450} (+/- 1 SD) of antibodies tested in triplicate is shown. The grey line indicates the background on the plate. The control antibodies are described in Table 2-35 (page 111).

The commercial anti-dsDNA and anti-insulin positive control antibodies bound their respective antigens under every blocking condition. However, ED38 and ei-JB40 only gave a positive result when the ELISA plate was blocked with 3% (w/v) BSA or 0.05% (v/v) Tween-20, 2 mM EDTA (Figure 3-12). This suggests that ED38 and ei-JB40 might not be polyreactive in the sense that they bind multiple unrelated antigens specifically. However they clearly show a distinct phenotype from the commercial positive controls and the negative control m-GO53. The cause of this phenotype is unclear and is therefore investigated in more detail in section 3.5.2 using antibodies discovered in the Dunn-Walters lab. In order to avoid confusion, this phenotype will henceforth be referred to as “promiscuous”.

3.5.2 Investigating the Promiscuous Antibody Phenotype

Five antibodies displaying a promiscuous phenotype were discovered in the Dunn-Walters lab by Tihomir Dodev (see Table 3-16 for details). These antibodies were used to further investigate the properties of the promiscuous phenotype and will be referred to as the "GF antibodies".

The ELISA shown in Figure 3-13 was conducted exactly as stated in Tiller et al. (2008) and shows that GF1, GF5, GF7, GF9 and GF16 all showed a positive result in wells coated with dsDNA, insulin, LPS, HEp-2 cell lysate and PBS (blank).

Table 3-16: Properties of the promiscuous GF antibodies. All chains are from the B cell Development dataset.[91]

Antibody ID		Chain ID	Donor	Cell Type	IGHV/IGKV Gene	IGHD/IGKD Gene	IGHJ/IGKJ Gene	CDR3 AA sequence
GF01	IgH	naih119_M_V3_103755	119	Naïve	IGHV3-30	IGHD2	IGHJ2	ARVVGSSKWDHAWYFDL
	IgK	nail119_IGKV3_UN_122382	119	Naïve	IGKV1-37	-	IGKJ5	QQYNNWPRT
GF05	IgH	naih140_M_V3_26308	140	Naïve	IGHV3-23	IGHD3	IGHJ4	AKASLVRYFDWLFNFDY
	IgK	nail119_IGKC_IGKV3_129323	119	Naïve	IGKV1-27	-	IGKJ3	QKYNSAPPFT
GF07	IgH	trah141_M_UN_42525	141	Transitional	IGHV3-33	IGHD3	IGHJ6	ARDMVLEWSYYYYGMDV
	IgK	tral162_IGKV1_UN_151849	162	Transitional	IGKV1D-12	-	IGKJ2	QQANSFPWT
GF09	IgH	trah149_M_V3_43529	149	Transitional	IGHV3-33	IGHD4	IGHJ6	ARGPHRYGDYGGYYYYYGMDV
	IgK	tral132_IGKV1_IGKC_124750	132	Transitional	IGKV1D-12	-	IGKJ5	QQANSFPIT
GF16	IgH	trah141_M_V3_68026	141	Transitional	IGHV3-33	IGHD3	IGHJ6	ARGTLDDFWSGYSLAGTINYPYYYGMDV
	IgK	tral162_IGKV1_UN_151849	162	Transitional	IGKV1D-12	-	IGKJ2	QQANSFPWT

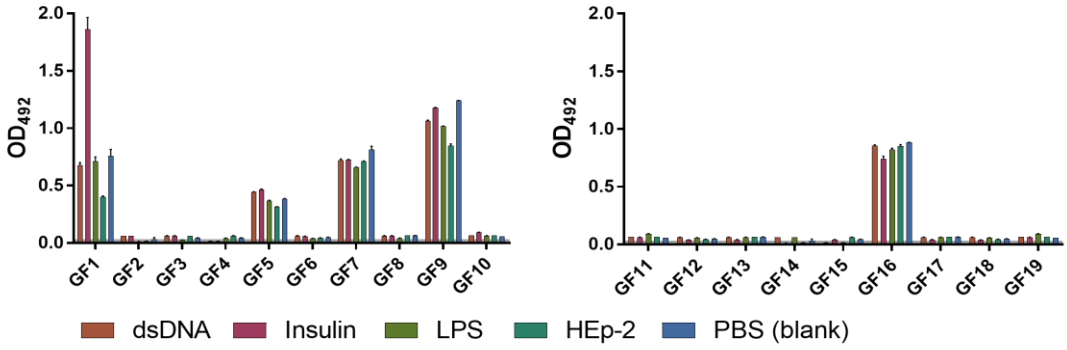


Figure 3-13: Discovery of the promiscuous GF antibodies using ELISA. Unpurified antibodies GF1, GF5, GF7, GF9 and GF16 showed a promiscuous phenotype when following Tiller et al. (2008) protocol (Tween-20 block). Each graph represents a separate ELISA plate.[91]

[91] Tihomir Dodev (KCL) discovered the GF antibodies in the B cell Development dataset and conducted the ELISA shown in Figure 3-13; this graph has been published in Laffy et al. (2016). The antibodies in Figure 3-13 were expressed by Tihomir Dodev using the method described in section 2.8 (page 103).

3.5.2.1 ELISA Conditions Affect the Promiscuous Phenotype

Figure 3-12 (page 165) shows that blocking conditions affect the visibility of the promiscuous phenotype. Therefore, a number of blocking and incubation conditions were tested to determine which resulted in the strongest promiscuous signal in solid-phase ELISA.

Three blocking conditions were tested:

1. 3% (w/v) milk powder in PBS
2. 3% (w/v) BSA in PBS
3. 0.05% (v/v) Tween-20, 2 mM EDTA in PBS

And two incubation conditions were tested:

1. 37°C, 200 rpm
2. Room temperature, static

The promiscuous phenotype was strongest when the ELISA plate was blocked with BSA and incubated at 37°C, 200 rpm (Figure 3-14). In this ELISA, many of the GF antibodies did not show evidence of binding in any of the conditions. Of those that did, the OD_{450} in the Tween-20 block was much lower than in Figure 3-13and they appear to bind insulin, in particular on the BSA-blocked wells; this was also a characteristic of ED38 (Figure 3-12, page 165).

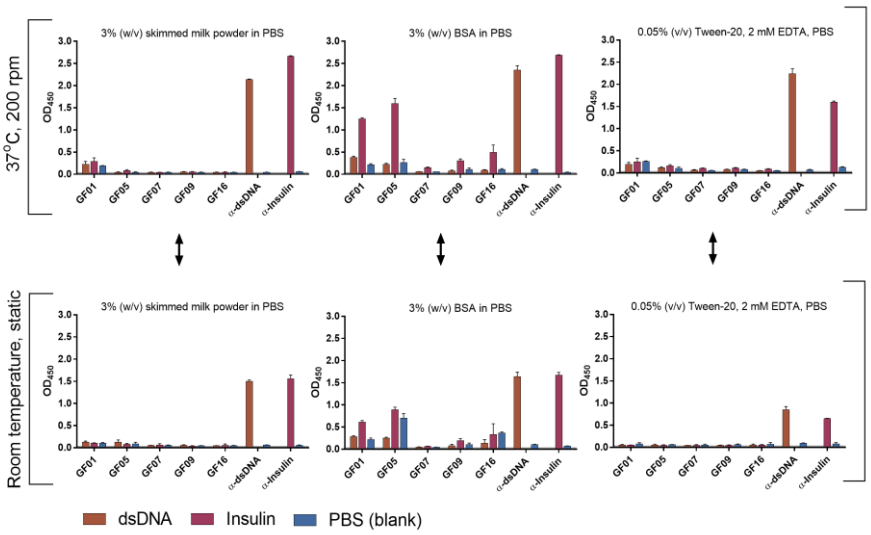


Figure 3-14: ELISA blocking conditions affect the promiscuous phenotype. Mean OD_{450} (+/- 1 SD) of antibodies tested in triplicate under three blocking conditions: 3% (w/v) skimmed milk powder in PBS, 3% (w/v) BSA in PBS and 0.05% (v/v) Tween-20, 2 mM EDTA in PBS and two incubation conditions: 37°C, 200 rpm and room temperature, static. Each graph represents a separate ELISA plate and the grey lines indicate the background on each plate.[92]

[92] The antibodies used in Figure 3-14 were expressed by the IgG team at MedImmune (Cambridge, United Kingdom) using the method described in section 2.7 (page 99). The control antibodies are given in Table 2-35 (page 111).

3.5.2.2 Promiscuous GF Antibodies do not bind Insulin in a Soluble ELISA

Unlike commercial positive control antibodies, the ELISA results for the promiscuous GF antibodies could be changed by altering the blocking and incubation conditions (Figure 3-14). This suggests that the promiscuous antibodies may not be interacting with the antigens coating the ELISA wells. To test this, a soluble ELISA was performed using streptavidin-coupled magnetic beads coated with biotinylated insulin (see Figure 3-15 for schematic and section 2.11 (page 112) for full materials and methods). As the magnetic beads are transferred to a new ELISA well before adding the HRP substrate, the colour will only develop if the antibodies are bound to the beads. The soluble ELISA was blocked with 3% (w/v) BSA and incubated at 37°C, 200 rpm.

The soluble ELISA showed no evidence that the promiscuous GF antibodies bind biotinylated insulin (Figure 3-16).

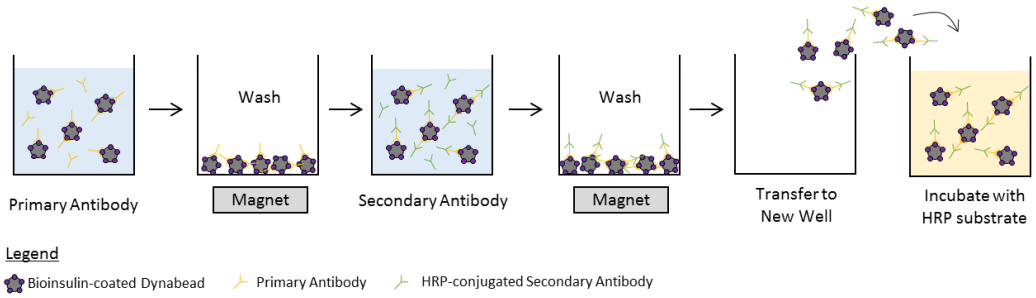


Figure 3-15: Soluble ELISA method. Streptavidin-coupled magnetic beads coated with biotinylated insulin are incubated with primary and secondary antibodies. Magnets are used during the wash steps. The bead-antibody complexes are transferred to a new well before the addition of the HRP-substrate.

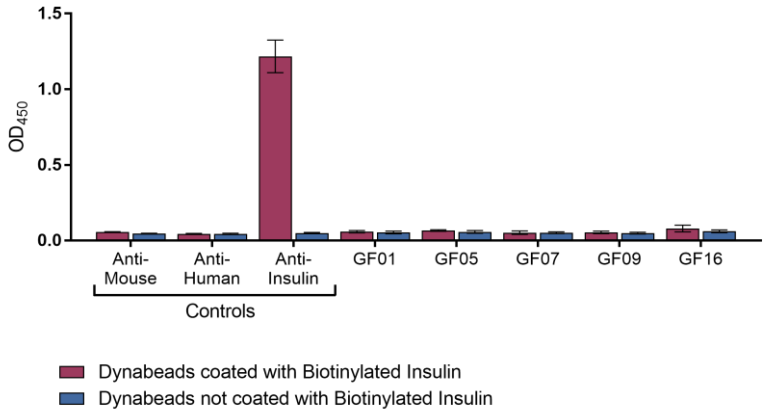


Figure 3-16: Soluble ELISA to test binding of GF antibodies to biotinylated insulin. Mean OD_{450} (+/- 1 SD) of antibodies tested in triplicate.[93]

[93] The GF antibodies used in Figure 3-16 were expressed by the IgG team at MedImmune (Cambridge, United Kingdom) using the method described in section 2.7 (page 99) and the control antibodies are detailed in Table 2-35 (page 111).

3.5.2.3 Promiscuous Antibodies are Autoreactive

The promiscuous GF antibodies appear to bind dsDNA on an ELISA plate under certain conditions (Figure 3-13, Figure 3-14). Immunocytochemistry against HEp-2 cells was used (as in section 3.4, page 153) to determine whether the GF antibodies show any nuclear (i.e. dsDNA) binding, and whether the promiscuous antibodies are autoreactive.

Only GF07 showed some evidence of nuclear binding; the other four antibodies did not bind the nucleus (see Figure 3-17 and Table 3-17). This suggests that positive result seen with dsDNA-coated ELISA wells (Figure 3-13 & Figure 3-14) was not dsDNA specificity. All five promiscuous GF antibodies showed cytoplasmic HEp-2 binding.

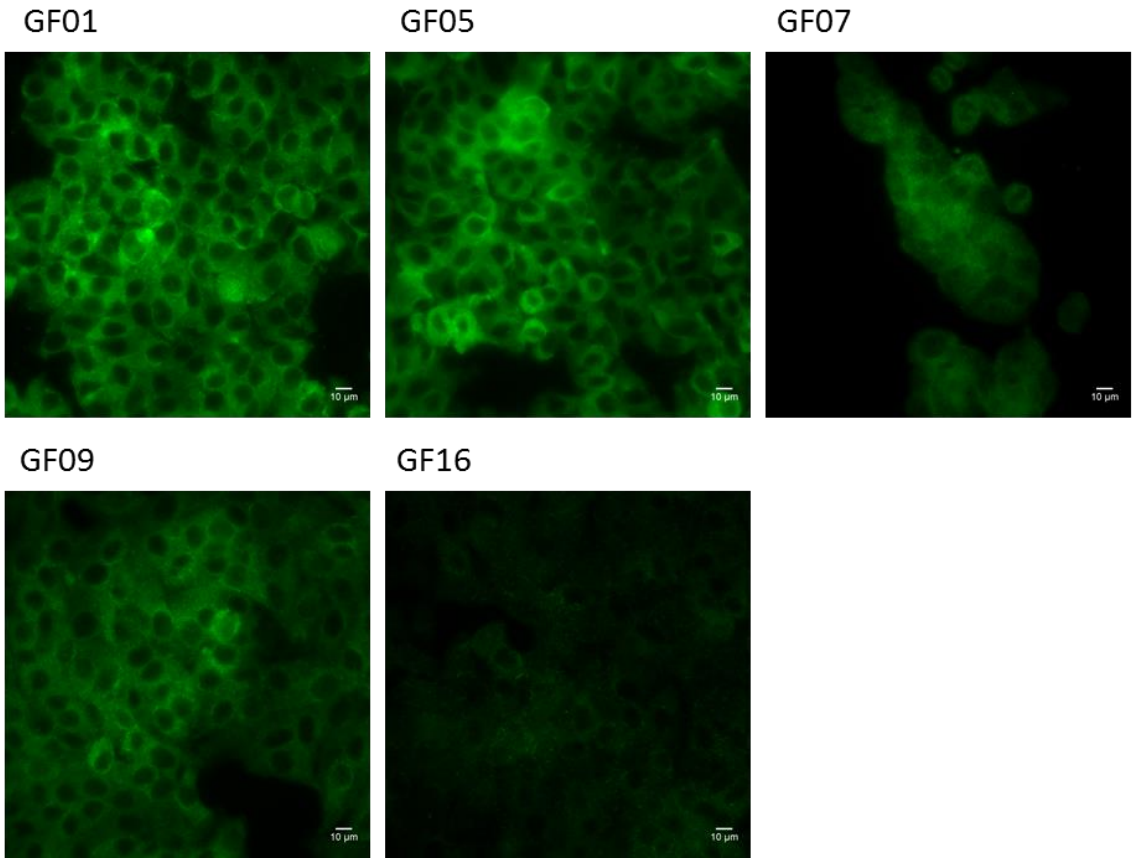


Figure 3-17: Immunocytochemistry to test the binding of the GF antibodies to HEp-2 cells. Despite apparently binding dsDNA on the solid-phase ELISA, these antibodies did not bind dsDNA in the nuclei of HEp-2 cells. All five GF antibodies bound the cytoplasm.

Table 3-17: HEp-2 binding specificities of GF antibodies. Refers to results shown in Figure 3-17.

Antibody ID	HEp-2 binding	Cytoplasmic binding	Nuclear binding
GF01	Strong	✓	×
GF05	Strong	✓	×
GF07	Strong	✓	✓
GF09	Strong	✓	×
GF16	Weak	✓	×

3.5.3 Investigating the Predicted Autoreactive Antibodies for Promiscuity

It was next investigated whether any of the antibodies cloned and tested for autoreactivity (sections 3.3 and 3.4 respectively) display a promiscuous phenotype.

3.5.3.1 Three Antibodies showed a Strong Promiscuous Phenotype

All of the antibodies which could be expressed out of the 61 cloned in section 3.3 were tested for the promiscuous phenotype using a solid-phase ELISA blocked with 3% (w/v) BSA in PBS and incubated at 37°C, 200 rpm. Each antibody was tested in triplicate against the following antigens:

1. dsDNA
2. Insulin
3. LPS
4. Blank wells (PBS only)

A number of antibodies showed a low-level promiscuous phenotype, however two antibodies (CT0053 and CT0054) showed a very strong promiscuous phenotype (Figure 3-18).

A subset of the antibodies investigated in Figure 3-18 were tested in a second ELISA using the same blocking and incubation conditions, but against a larger panel of antigens. In this ELISA each antibody was tested in triplicate against the following antigens:

1. dsDNA
2. Insulin
3. GIF (gastric intrinsic factor)
4. LPS
5. MPO (myeloperoxidase)
6. Blank wells (PBS only)

Again, CT0053 and CT0054 showed a very strong promiscuous phenotype, giving a positive signal in all the antigen-coated wells and the PBS-only wells (Figure 3-19). However, in this ELISA, CT0037 also showed a very strong promiscuous phenotype, whereas in Figure 3-18 it had appeared only weakly promiscuous.

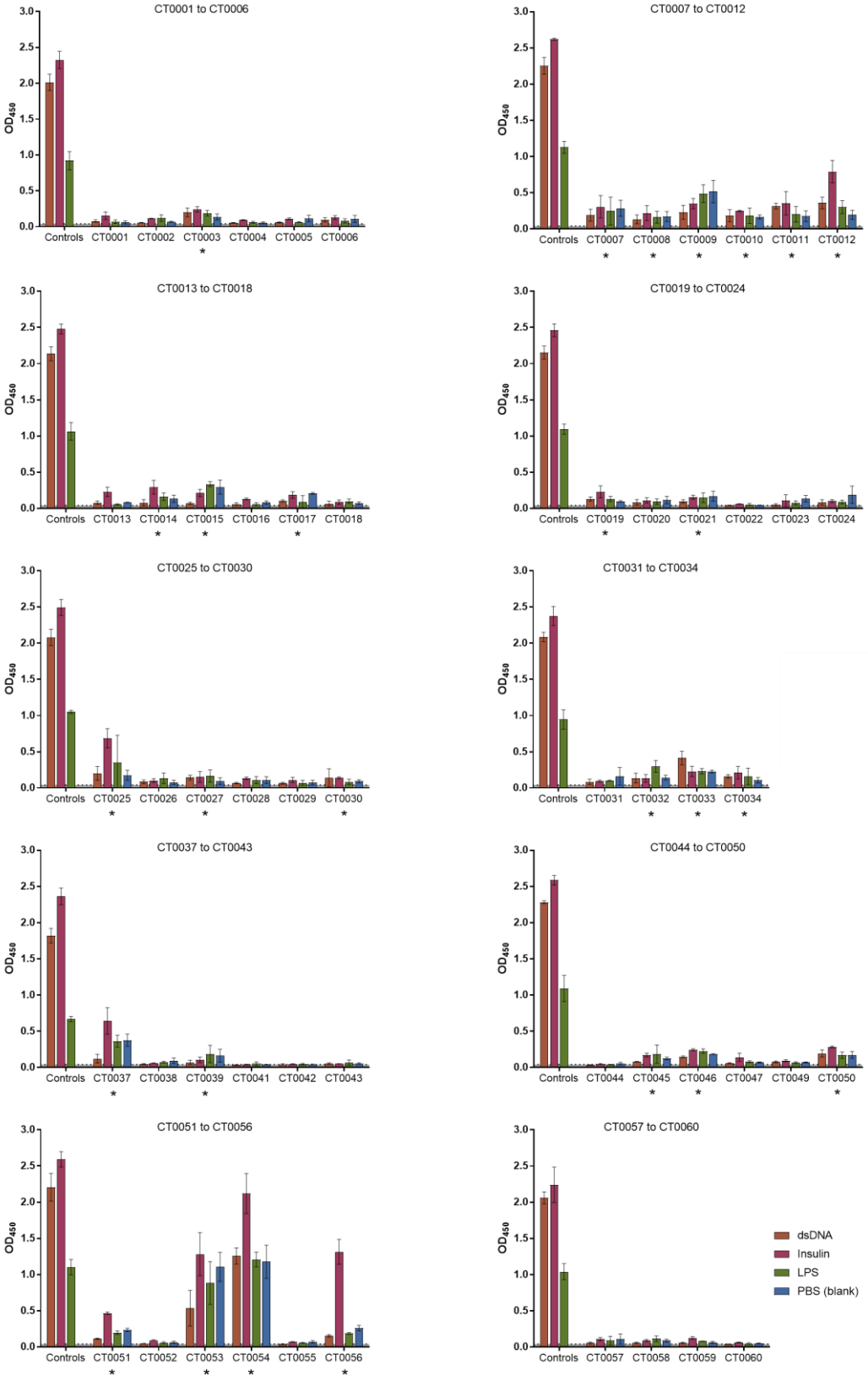


Figure 3-18: ELISA testing the *Forbidden* antibodies for the promiscuous phenotype. Mean OD_{450} (+/- 1 SD) of each antibody tested in triplicate. All the ELISA plates were blocked with 3% (w/v) BSA in PBS and the primary antibodies were incubated at 37°C, 200 rpm. Asterisks indicate promiscuous antibodies. Each graph represents a separate ELISA plate. The grey line indicates the background on each plate. The control antibodies are described in Table 2-35 (page 111).

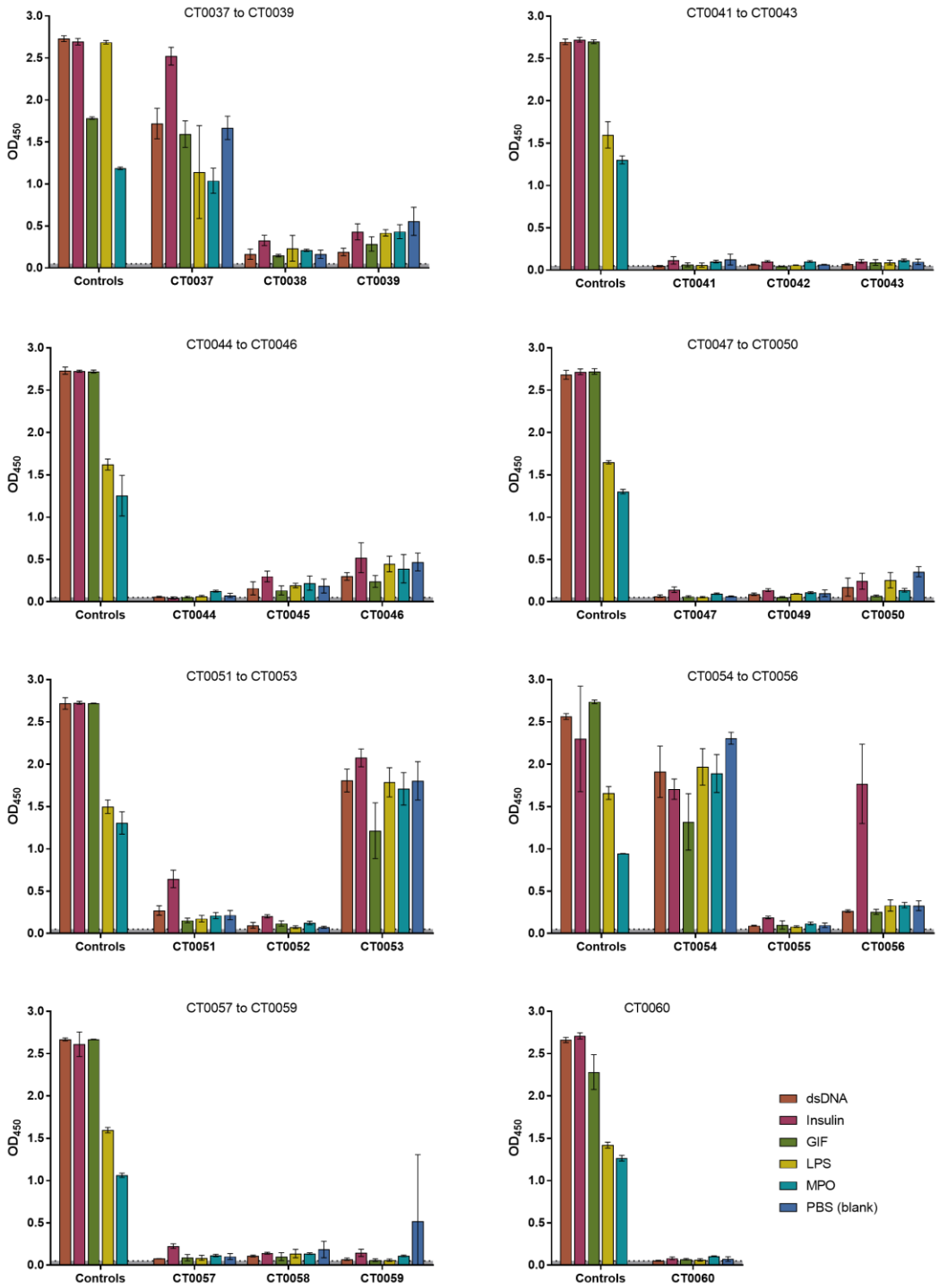


Figure 3-19: ELISA testing antibodies against further antigens. OD_{450} (+/- 1 SD) of each antibody tested in triplicate. All the ELISA plates were blocked with 3% (w/v) BSA in PBS and the primary antibodies were incubated at 37°C, 200 rpm (as in Figure 3-18). Each graph represents a separate ELISA plate. The grey line indicates the background on each plate. The control antibodies are described in Table 2-35 (page 111).

3.5.3.2 Three Promiscuous Antibodies do not bind Insulin in a Soluble ELISA

The three antibodies which showed a promiscuous phenotype in Figure 3-19 (CT0037, CT0053 and CT0054) were tested in a soluble ELISA against streptavidin-coupled magnetic beads coated with biotinylated insulin (as in section 3.5.2.2, page 168). None of the three antibodies showed evidence of binding to the biotinylated insulin in a soluble ELISA format (Figure 3-20).

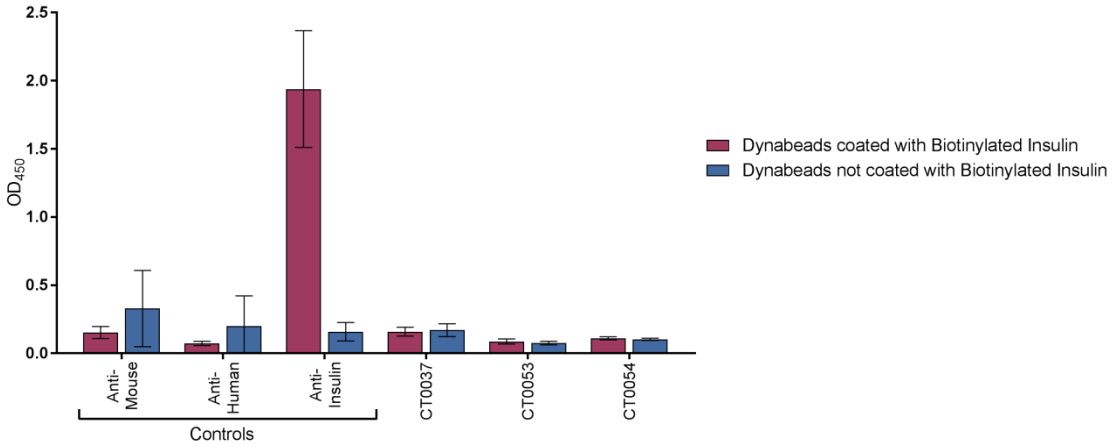


Figure 3-20: Three promiscuous antibodies do not bind insulin in soluble ELISA. Mean OD_{450} (+/- 1 SD) of each antibody. CT0037, CT0053 and CT0054 were tested in triplicate over two ELISA plates, each containing a set of controls. In this figure, the controls from both plates have been pooled and therefore the mean (+/- 1 SD) is based on six data points. The control antibodies are detailed in Table 2-35 (page 111).Three Promiscuous Antibodies show Aggregation

It was noticed that the three promiscuous antibodies (CT0037, CT0053 and CT0054) were also the three antibodies with the highest percentage aggregation in the HP-SEC traces provided by the MedImmune IgG team when the antibodies were purified in PBS (Figure 3-21). See Appendix Figure A-4 (page 283) for all antibody HP-SEC traces.

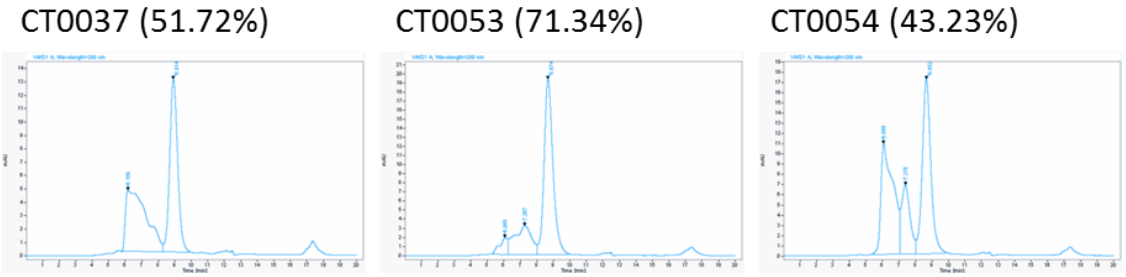


Figure 3-21: HP-SEC trace for CT0037, CT0053 and CT0054. The purified CT0037, CT0053 and CT0054 comprised 51.72%, 71.34% and 43.23% monomers respectively. The large peak at 9 min represents IgG1 monomers. The peaks to the left of 9 min (shorter time periods) are aggregated IgG. Any peaks to the right of 9 min (longer timer periods) are fragmented IgG1.

3.5.3.4 Three Promiscuous Antibodies show Unusual Thermal Stability Patterns

A thermal shift assay was used to determine the melting temperature (Tm) of each antibody expressed. For full materials and methods, see section 2.12 (page 113). Briefly, each purified antibody is mixed in an aqueous solution with SYPRO Orange (a fluorophore which is quenched by water). The mixture is gradually heated in a real-time PCR machine and the fluorescence is measured regularly. Once the Tm is reached and the protein begins to unfold, hydrophobic pockets are revealed and the SYPRO Orange sequestered in these pockets fluoresces. Eventually the protein disintegrates and the fluorophore is quenched (see Figure 3-22A for a schematic).

A thermal shift assay was performed for every antibody expressed. Most antibodies started fluorescing between 60°C and 70°C. However, CT0037, CT0053 and CT0054 all showed an unusual curve in which there was already fluorescence at 20°C (Figure 3-22B), particularly CT0037 and CT0054, which were the most aggregated in Figure 3-21.

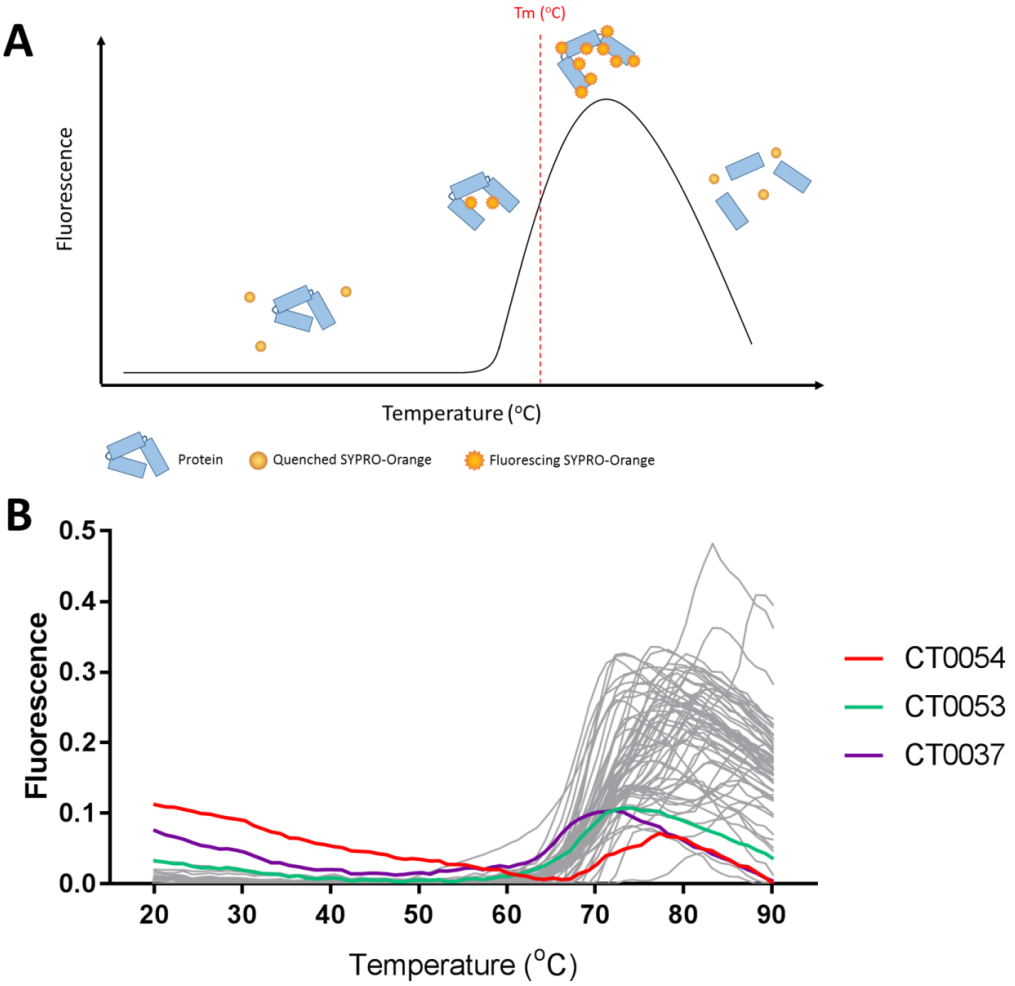


Figure 3-22: Thermal shift assay using SYPRO-Orange. (**A**) Principal of a thermal shift assay. (**B**) Thermal shift assay for all antibodies expressed. Each antibody was tested in triplicate and the line represents the mean of each reading. Antibodies CT0037, CT0053 and CT0054 are highlighted in blue, green and red respectively.

3.5.4 Promiscuous Antibody Phenotype may be linked to Autoreactivity

22 of the 27 antibodies with a promiscuous phenotype (starred antibodies in Figure 3-18, page 171) were autoreactive (as defined by HEp-2 binding), whereas only 17 of the 29 antibodies which did not have the promiscuous phenotype were autoreactive (Figure 3-23A). However, a two-sided Fisher's exact test did not find this difference significant ($p = 0.0843$).

When looking at where in the cell autoreactive antibodies bound (Figure 3-23B), there was no significant difference between the promiscuous antibodies and the not promiscuous antibodies (two-sided Fisher's exact test $p = 0.9298$).

However, it was observed that the three antibodies which showed a very strong promiscuous phenotype (CT0037, CT0053 and CT0054) (Figure 3-19, page 172) also showed very strong HEp-2 cytoplasmic binding (Figure 3-6, page 158, and Figure 3-8, page 160).

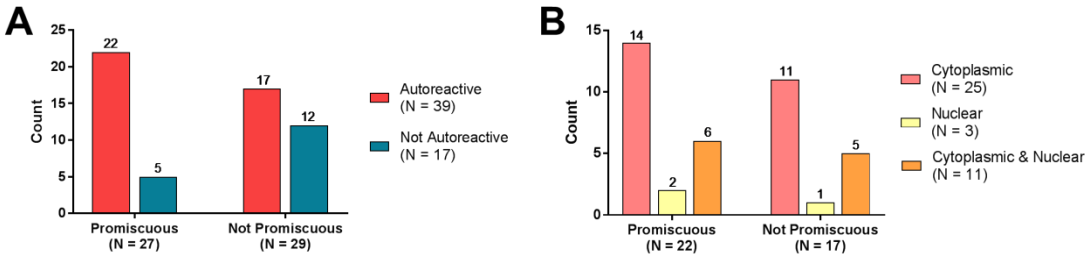


Figure 3-23: Promiscuous antibodies are more likely to be autoreactive. The count of each group is displayed above the relevant bar.

3.6 Selecting for Polyreactive Antibodies using Phage Display

In section 3.5, it was found that some antibodies displaying an apparent affinity for multiple antigens in certain solid-phase ELISA formats may not be binding the antigens with which the wells are coated. We named this phenotype "promiscuous". This suggested that polyreactive antibodies which bind multiple unrelated antigens specifically may not be as common as previously thought (Wardemann et al., 2003). To investigate whether we could find polyreactive antibodies in the human repertoire, we used a phage display selection cascade. The principle of phage display is described in section 1.12.2.1 (page 68). Briefly, a phage display library is created by inserting scFv sequences into the bacteriophage genome, causing the scFv to be expressed on the phage surface. Panning selections are performed by incubating the phage library with the antigen of interest; those expressing antigen-specific scFv will bind the antigen and the non-binders will be washed away. The prevalence of antigen-specific scFv increases in the library with every round of selection. We hypothesised that, if polyreactive antibodies are common in the human B cell repertoire, it should be possible to isolate them by selecting against a different antigen at each round of selection as only the antibodies which bound all of the antigens (i.e. polyreactive) would remain. The phage display library used was the MedImmune in-house BMVtrp library. This library is an antigen-naïve human scFv library constructed from peripheral blood (15 donors) tonsil (4 donors) and bone marrow (24 donors) B cell mRNA (Vaughan et al., 1996).

The BMVtrp library was selected against the following unrelated autoantigens (Figure 3-24):

1. HEp-2 cell lysate
2. dsDNA
3. MPO (myeloperoxidase)
4. ApoH (apolipoprotein H)

To prevent the selection of promiscuous antibodies (as defined in section 3.5.1, page 163), the library and the wells were blocked with 3% (w/v) milk powder in PBS. This blocking buffer has been shown to inhibit the binding of promiscuous antibodies (Figure 3-12, page 165). Additionally, before being incubated with each antigen, the library was first incubated in a blank ELISA well which had been blocked with 3% (w/v) milk powder in PBS. The aim of this was to try to remove any scFv which were just binding the plastic well or protein block. To mitigate the risk of the selection failing at one of the antigens, the selection cascade was split into multiple arms using the antigens in different orders (Figure 3-24). The selection was monitored by measuring the input and output titres at each round of selection. The diversity was measured after each round of selection by sequencing 88 output colonies. Of each batch

of 88 output colonies, each unique scFv was tested in an anti-M13 phage ELISA, blocked with 3% (w/v) milk powder in PBS, to determine if the antibodies bound multiple antigens (dsDNA, MPO and blank (PBS only) wells). The successful expression of M13 phage was confirmed for every ELISA well to ensure that a negative result was not due to lack of phage expression. Commercial anti-dsDNA and anti-MPO positive controls confirmed the presence of the antigens in the ELISA wells. For full materials and methods, see section 2.13 (page 114).

The antigen ELISAs showed that none of the scFv isolated in any of the rounds of selection shown in Figure 3-24 bound specifically to both dsDNA and MPO, suggesting that the selection had not isolated polyreactive antibodies.

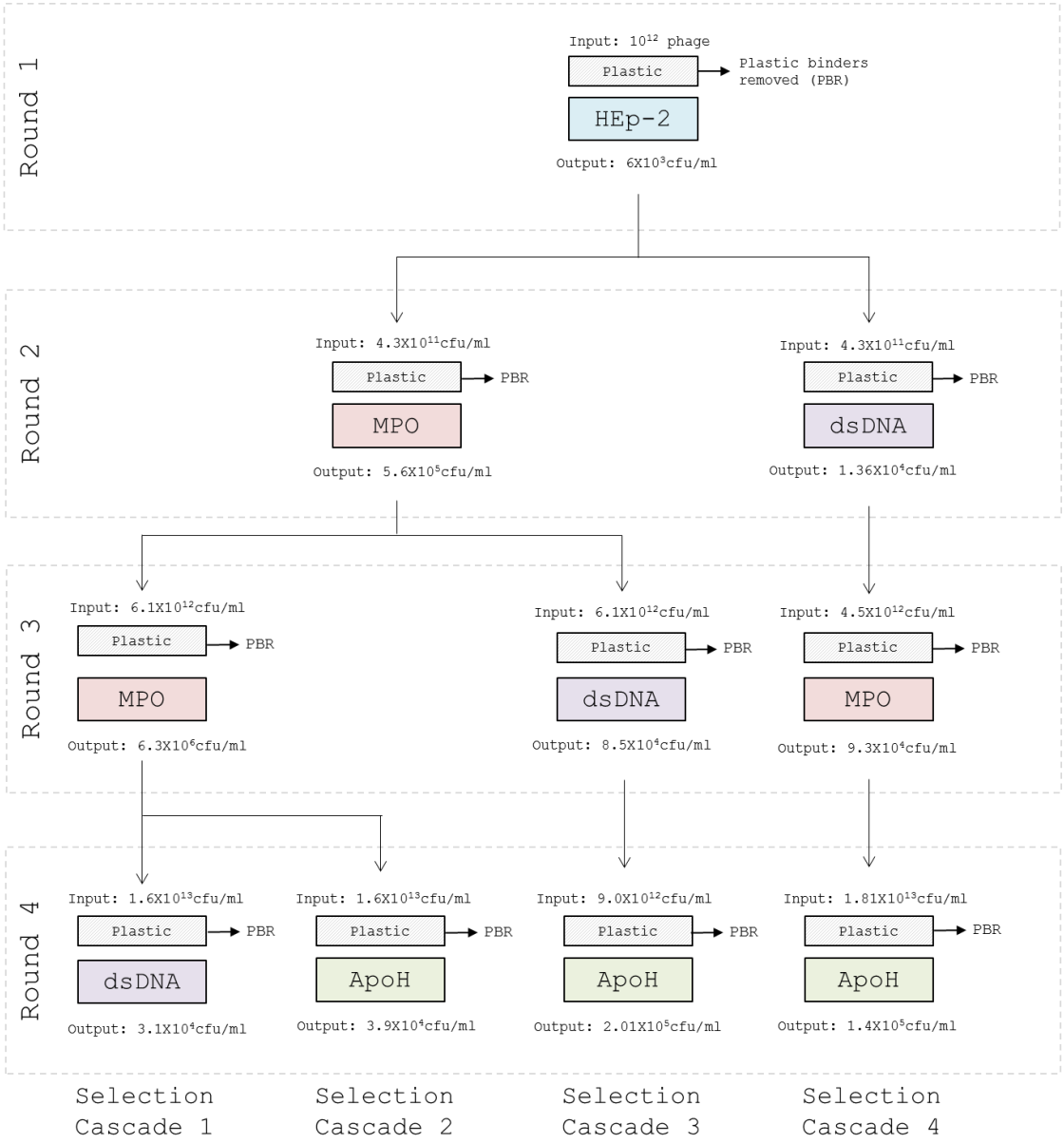


Figure 3-24: Phage display selection cascade against multiple unrelated autoantigens. The BMVtrp library was selected against multiple unrelated autoantigens with the aim of isolating polyreactive scFv. The plates were blocked with milk and plastic-binders were de-selected at each round.

3.7 Discussion

3.7.1 Transitional Populations in the B cell Development Pathway

The original hypothesis of this chapter was that autoreactive B cells are selected against at the central and peripheral tolerance checkpoints and therefore any variable region characteristics associated with autoreactivity would reduce in frequency as the B cells matured and any preferred characteristic would increase in frequency as B cells matured. One of our starting assumptions was that B cells progress linearly from pre-B >> immature >> transitional >> naïve. Therefore, we expected the frequency of each variable region property to show one of three patterns as the B cells matured through the four cell repertoires:

1. No change in frequency, indicating no selection pressure
2. Progressive increase in frequency, indicating that the property was positively selected
3. Progressive decrease in frequency, indicating that the property was selected against

However, a surprising finding in the B cell Development dataset analysis was that the transitional B cell variable region properties did not always fit the trajectory of gene loss/gain or CDR-H3 property increase/decrease expected if the whole B cell population progresses linearly from pre-B to naïve. For example, IGHD3 and IGHJ4 are used significantly more and less frequently respectively in the transitional repertoire when compared to the other three cell types (Figure 3-1, page 140). Furthermore, the frequency of IGHJ6 is significantly higher in the transitional repertoire compared to pre-B and immature, but the frequency in the naïve repertoire is significantly lower than in the immature repertoire. Some CDR-H3 properties also showed this trait; the mean CDR-H3 aliphatic index was significantly lower in the transitional repertoire and there was also a significantly higher frequency of CDR-H3 with a high isoelectric point in the transitional repertoire (Figure 3-2B&D respectively, page 142).

The transitional repertoire can be split into three subsets: T1 ($CD38^{+++}CD24^{hi}CD10^{++}IgD^{lo/-}$), T2 ($CD38^{++}CD24^{hi}CD10^{+}IgD^{+}$) and T3 ($CD38^{+}CD24^{+}IgD^{+}ABCB1^{-}$) (Martin et al., 2016) and flow cytometry data suggests that the transitional cells form a continuum between the immature and naïve repertoires (Palanichamy et al., 2009). However, if B cells do mature in a linear fashion from pre-B >> immature >> transitional >> naïve, it would waste resources to promote a property in one cell population and then select against it in the subsequent one (or vice versa). Therefore, our findings suggest that the progression from pre-B to naïve may not be linear and the whole of the transitional repertoire may not mature into naïve B cells.

The role of the transitional repertoire is still not completely understood, however it has been found that both T1 and T2 subsets are present in the bone marrow and peripheral blood of

healthy adults, suggesting that both subsets can leave the bone marrow and both may have independent roles in the periphery (Agrawal et al., 2013). This raises questions about the linearity of the transition from immature >> T1 >> T2 >> T3 >> naïve. Our work supports the hypothesis that the transitional B cell repertoire contains a large population of cells which are not part of the linear development pathway from pre-B >> immature >> transitional >> naïve. We hypothesise that some of the B cells in the transitional repertoire are destined to form other B cell repertoires and are thus subject to different selection pressures (Martin et al., 2016).

Despite these surprising findings, there is evidence that central and peripheral tolerance do remove autoreactive B cells from the repertoire (Wardemann et al., 2003) and therefore we still analysed the B cell Development dataset to identify genes and CDR-H3 properties which may be associated with autoreactivity. However, when selecting heavy chain properties to investigate, emphasis was placed on the difference in frequency between the pre-B/immature repertoire and the naïve repertoire rather than the frequency in the transitional population as not all of the transitional population may be subject to the same tolerance pressures.

3.7.2 Selection of Heavy Chain Variable Region Properties as B cells Mature

To identify which heavy chain properties may be associated with autoreactivity, we hypothesised that properties which decreased in frequency as B cells matured may do so because they are selected out of the repertoire because they tend to be present in autoreactive antibodies. Conversely, we hypothesised that properties which increased in frequency as B cells matured may be beneficial.

The only IGHV family which significantly changed in frequency of usage as B cells matured was IGHV3. IGHV3 was significantly less frequent in the naïve repertoire compared to all three prior repertoires (Figure 3-1A, page 140). This appeared to be driven by decreases in the frequency of IGHV3-30, IGHV3-33, IGHV3-15 and IGHV3-7 (Figure 3-1D). Although the IGHV1 family was more frequent on average in the naïve repertoire (Figure 3-1A), the difference was not statistically significant. However, IGHV1-18 and IGHV1-69 genes were significantly more common in the naïve repertoire. IGHV1-69 is an important antiviral gene due to the presence, in many alleles, of a phenylalanine at position 54 (Kabat scheme) which is at the apex of the CDR-H2 loop (Avnir et al., 2016). It has been noted previously that the IGHV3 family is less frequent and the IGHV1 family more frequent in switched memory populations compared to naïve populations (Wu et al., 2010), indicating that these families commonly highlight differences between cell repertoires.

IGHV4-34 is heavily associated with autoimmune disease such as SLE, RA and MS (Tipton et al., 2015, Doorenspleet et al., 2014, Baranzini et al., 1999). We found that IGHV4-34 is significantly less common in the immature and transitional repertoires compared to the pre-B repertoire, suggesting that it may be selected against early in development. However, surprisingly, the frequency of IGHV4-34 significantly increases again in the naïve repertoire (Figure 3-1D), suggesting that this gene may, in some circumstances, be advantageous.

The only IGHD family which significantly decreased in frequency as B cells matured was IGHD2. The IGHD2 family was significantly less common in the transitional and naïve repertoires compared to pre-B and immature (Figure 3-1B). This appeared to be driven by significant decreases in IGHD2-15 and IGHD2-2 usage (Figure 3-1E).

IGHJ6 was significantly less common in the naïve repertoire compared to the immature repertoire, suggesting that it may be selected against (Figure 3-1C). IGHJ3 is significantly more frequent in the naïve repertoire, suggesting that it is preferred.

The reason IGHD2-2, IGHD2-15 and IGHJ6 appear to be selected out of the repertoire as B cells mature may be due to the fact that they are very commonly found in antibodies with long CDR-H3. Long CDR-H3 are associated with autoreactivity (Wardemann et al., 2003). IGHD2-2 and IGHD2-15 are among the longest IGHD genes and IGHJ6 is the longest IGHJ gene. IGHJ6 is present in over 75% of the longest CDR-H3 and IGHJ6 in combination with either IGHD2 or IGHD3 genes is seen in 58.8% of long CDR-H3 (Briney et al., 2012). However, IGHD3-3 is also very common in long CDR-H3 but this gene did not show evidence of being selected against, suggesting that the IGHD2-2 and IGHD2-15 genes may have additional properties which are selected against at tolerance. The IGHD2 family is more frequent in non-productive sequences than productive rearrangements (again indicating that the family is selected out of the repertoire), but has not been linked to autoreactive antibodies (Volpe & Kepler, 2009, Volpe & Kepler, 2008).

The mean CDR-H3 length was shorter in the naïve repertoire compared to the other three repertoires (Figure 3-2A, page 142), suggesting that long CDR-H3 are selected against. Long CDR-H3 are often associated with autoreactivity, however the reason for this association is not well understood. It is speculated that it could be due to increased N-nucleotide addition and so the addition of non-germline-encoded amino acids which have not been subject to millennia of natural selection. There is a positive correlation between CDR-H3 length and N-nucleotide and P-nucleotide addition (Briney et al., 2012). Although potentially dangerous due to the risk of autoreactivity, long CDR-H3 loops are important for fighting off certain infections, particularly HIV, malaria and trypanosomes (Pancera et al., 2010, Henderson et al., 2007, Stijlemans et al.,

2004). For example, two potent HIV broadly-neutralising antibodies (PG9 and PG16) both have CDR-H3 loops 28 amino acids long (Pancera et al., 2010). Long CDR-H3 can be useful because they can reach epitopes otherwise obscured in deep crevices or by heavy glycosylation.

The mean hydrophobicity (GRAVY index) was significantly lower in the transitional and naïve repertoires compared to the pre-B and immature repertoires (Figure 3-2E), indicating that hydrophobic CDR-H3 are selected against. The mean Boman index of CDR-H3 was significantly higher in the naïve repertoire compared to the three preceding repertoires (Figure 3-2C). The Boman index is a measure of antibacterial (anti-membrane) activity (Boman, 2003) and was designed as an index to specifically measure propensity to bind bacterial membranes but not other proteins. Therefore, the promotion of antibodies with a high Boman index might reflect the positive selection of antibodies which bind bacterial membranes but not self-antigens.

Almost no significant differences were seen between the pre-B and immature repertoires: only IGHV3-15 and IGHV4-34 were used at different frequencies and the isoelectric point and instability index of the CDR-H3 was significantly different. This lack of difference between the pre-B and immature repertoires supports the hypothesis that no selection takes place between these two repertoires and that pre-B proliferation is driven by its ability to recognise self-antigen, indicating that it has functional BCRs (Köhler et al., 2008).

3.7.3 Long, Hydrophobic and Aromatic CDR-H3 Linked to Cytoplasmic Binding

The first set of heavy chains selected to test for autoreactivity possessed all of the properties mentioned above which had been identified as forbidden, and were therefore named the *Forbidden* antibodies. These heavy chains all used IGHV3-15/3-30/3-33, IGHD2-2 and IGHJ6 genes and they all had exceptionally long, hydrophobic CDR-H3 with low CDR-H3 Boman indexes. In order to investigate the roles of the heavy and light chains independently, each heavy chain was paired with three public light chains. Of the 12 heavy chains investigated, six (IgH4, 5, 8, 9, 10 and 11) bound the HEp-2 cytoplasm when paired with all three light chains (Figure 3-5, page 156). Only two heavy chains (IgH3 and 6) did not bind the HEp-2 cytoplasm when paired with any of the light chains. IgH3 was the only heavy chain derived from a naïve B cell, suggesting that this cell had passed the tolerance checkpoints and, despite its risky properties, is not autoreactive. The fact that IGH1, 2 and 7 show variable autoreactivity depending on the light chain with which they are paired suggests that these heavy chains are prone to autoreactivity but can be rescued by receptor editing. All three of these antibodies bound HEp-2 cells when in combination with one or both of the kappa light chains. This supports the theory that lambda light chains are better at rescuing autoreactive heavy chains than kappa (Wardemann et al., 2004). IgH12 could not be expressed with two of the three light

chains. This suggests that some heavy-light chain combinations do not work either because they do not pair well or because the resulting antibody is toxic to the cell. Nine out of 12 *Forbidden* heavy chains did show HEp-2 binding, indicating that these characteristics are associated with autoreactivity. However, a more robust experiment would have been to compare the *Forbidden* heavy chains with a set of randomly selected heavy chains and paired with the same three light chains to see if autoreactivity was more common in the *Forbidden* heavy chains. Furthermore, the 12 *Forbidden* heavy chains had such similar properties, there were no clear differences between those which did bind the HEp-2 cells and those that did not. This suggests that to get more accurate predictors of autoreactivity, one should probably look beyond the amino acid sequence and examine the structure.

Further evidence that antibodies with long, hydrophobic CDR-H3 are associated with autoreactivity is that three out of the four *Long Hydrophobic* antibodies which could be expressed bound HEp-2 cells (CT0050, CT0053, CT0054). CT0053 and CT0054 showed strong cytoplasmic binding, whereas CT0050 showed weak nuclear binding (Figure 3-6, page 158). CT0050 has a string of valines in the CDR-H3. Valines are small, non-polar and aromatic (Table 2-20 page 95). However, this string of valines is associated with the use of IGHD2 genes and all of the *Forbidden* antibodies used IGHD2-2 but they did not all show signs of nuclear binding. This suggests that it may be the long, hydrophobic CDR-H3 which is mainly responsible for the autoreactivity shown by these antibodies, rather than the VDJ genes as the VDJ gene usage was not considered when selecting the *Long Hydrophobic* heavy chains.

In the literature (Droupadi, 1994) and from experiments in the Dunn-Walters lab, aromatic antibodies show some association with autoreactivity/polyreactivity. Therefore, heavy chains with a CDR-H3 with a very high aromatic amino acid content were investigated for autoreactivity. Like the antibodies with long, hydrophobic CDR-H3 loops, the *Aromatic* antibodies showed a propensity to cytoplasmic binding. Three of the four *Aromatic* antibodies that could be expressed showed cytoplasmic binding (CT0037, CT0047, CT0060). CT0037 in particular showed very strong cytoplasmic binding (Figure 3-8, page 160).

3.7.4 Unusual CDR-H3 Isoelectric Points are Associated with Nuclear Binding

The analysis of the heavy chain B cell Development dataset showed that the frequency of antibodies in all four cell repertoires with CDR-H3 loops with an isoelectric point (pI) between 5.0-6.0 (net negative charge at neutral pH) or 7.0-8.5 (net neutral to positive charge at neutral pH) was very low (Figure 3-2D, page 142). This suggests that these properties may be disadvantageous and so very rarely occur in the repertoire, potentially because they are heavily associated with autoreactivity. Two groups of five antibodies with CDR-H3 loops with pI

values within these boundaries (*pI 5.0-6.0* and *pI 7.0-8.5*) were cloned and tested for autoreactivity.

Surprisingly, both groups of antibodies showed a tendency towards strong nuclear binding. Three of the *pI 5.0-6.0* antibodies showed nuclear binding (CT0039, CT0043, CT0049), two of which showed exclusively nuclear binding (CT0039 and CT0043) (Figure 3-9, page 161). Four of the five *pI 7.0-8.5* antibodies showed cytoplasmic binding (CT0046, CT0051, CT0055, CT0056), and three of these also showed nuclear binding (CT0046, CT0051, CT0056) (Figure 3-10, page 162). This suggests that these isoelectric points may be underused in the repertoire because they are prone to nuclear binding. Anti-nuclear antibodies are associated with various autoimmune disease, particularly SLE (Arbuckle et al., 2003), and alternating CDR-H3 charge patterns are associated with anti-nuclear antibodies in lupus mice (Liang et al., 2004).

3.7.5 Preferred Characteristics did not Guarantee Lack of Autoreactivity

Our hypothesis was that antibodies with properties which are selected against at tolerance would have a high probability of being autoreactive, but conversely, antibodies with properties which are promoted at the tolerance checkpoints would have a low chance of being autoreactive. We selected five *Preferred* heavy chains which possessed all the characteristics which significantly increased as B cells matured.

Surprisingly, of the *Preferred* antibodies, two showed both nuclear and cytoplasmic binding (CT0058, CT0059), however the binding was weak (Figure 3-7, page 159). This suggests that either just possessing preferred characteristics is not enough to guarantee no autoreactivity or certain anti-nuclear antibodies are not harmful. Both nuclear-binding heavy chains were from transitional cells indicating that they had passed central tolerance.

It is known that autoreactive antibodies do escape central tolerance and are found in the peripheral blood (Wardemann et al., 2003), and unless a complementary autoreactive T cell is also present, the B cell should not be activated (Parham, 2009). However, it was surprising that two autoreactive *Preferred* antibodies showed nuclear binding as this phenotype is heavily associated with SLE (Arbuckle et al., 2003).

However, it is important to remember that these heavy chains are all paired with the same light chain which is almost certainly not their native pair. It is possible that these heavy chains would not have shown autoreactivity if paired with their native light chains.

3.7.6 Antibody Polyreactivity may be Promiscuity

The work in this chapter was initially based on work in the landmark Wardemann et al. (2003) Science paper where it was shown that autoreactivity and polyreactivity are closely linked and, like autoreactive antibodies, polyreactive antibodies are removed from the repertoire at the tolerance checkpoints. They showed that polyreactive antibodies are very prevalent in the early immature B cell repertoire (55.2%) but approximately 90% are removed at the central tolerance checkpoint leaving 6.9% and 7.4% polyreactive antibodies in the immature and new emigrant B cell populations respectively. Peripheral tolerance then removes a further 42% of polyreactive B cells, leaving a frequency of approximately 4.3% in the naïve mature repertoire. This indicates that most polyreactive antibodies are not advantageous in the mature naïve B cell population, potentially because they may lead to autoimmune disease. Many papers have followed on from this 2003 study, each defining a polyreactive antibody as an antibody which binds multiple unrelated antigens (usually dsDNA, ssDNA, insulin and LPS) in the solid-phase ELISA protocol set out in detail in Tiller et al. (2008). In almost all of these papers ED38, ei-JB40 and m-GO53 are used as high-polyreactive, low-polyreactive and negative controls respectively. Studies using this definition of polyreactivity have found that the frequency of polyreactive antibodies differs dramatically in different B cell repertoires, for example, the frequency of polyreactive antibodies is significantly higher in IgA+ and IgG+ intestinal plasmablasts (Benckert et al., 2011, Berkowska et al., 2015) indicating that the polyreactive phenotype may play a role in maintaining mucosal immunity. Polyreactivity is less frequent in the IgM+ memory B cell compartment than the mature naïve repertoire (Tsuiji et al., 2006), but it is more common in the IgG+ memory B cell compartment due to somatic hypermutation (Tiller et al., 2007). This indicates that there may be an additional tolerance checkpoint between the mature naïve repertoire and the IgM+ memory B cell repertoire and there could be an additional tolerance checkpoint between the IgG+ memory repertoire and the antibody-secreting plasma cell repertoire. Polyreactivity is also more common among an unusual group of peripheral blood B cells which co-express normal light chains and surrogate light chains (V-preB+L+) (Meffre et al., 2004).

Polyreactivity (as defined in Tiller et al. (2008)) is also more common in patients with autoimmune diseases such as rheumatoid arthritis (Samuels et al., 2005) and both active systemic lupus erythematosus (SLE) and SLE in remission (Yurasov et al., 2005, Yurasov et al., 2006). Polyreactivity is more common in patients with various defects in B cell development such as those with X-linked agammaglobulinemia (defective Bruton's tyrosine kinase) (Ng et al., 2004), people with chronic lymphocytic leukaemia (particularly the non-mutated clones) (Hervé et al., 2005), CD40L-deficient patients (Hervé et al., 2007) and AID-deficient patients

(Meyers et al., 2011, Cantaert et al., 2015). Interestingly, antibodies which bind HIV gp140 and also antibodies which bind the influenza hemagglutinin (HA) stalk frequently show polyreactivity (Mouquet et al., 2010, Andrews et al., 2015) indicating that polyreactivity may be advantageous in protection against viral infections such as HIV and influenza.

The polyreactive phenotype described in all of the papers mentioned above is clearly biologically very important and can be advantageous in certain circumstances, but must be tightly controlled by various checkpoints due to the risk of autoimmune disease. We therefore wanted to test our panel of cloned antibodies (section 3.3, page 144) for polyreactivity using the same protocol. However, when we tried to replicate the ELISA described in these papers using ED38, ei-JB40 and m-GO53, we found that, when the plate was blocked with the protein-free blocking buffer described in Tiller et al. (2008), ED38, ei-JB40 and m-GO53 gave the same signal regardless of whether the ELISA wells were coated with antigen (Figure 3-11, page 164). Additionally, the binding pattern of these antibodies can be altered by changing the blocking conditions, whereas the binding of commercial antigen-specific antibodies remains unchanged (Figure 3-12, page 165). These results suggested that the polyreactive phenotype displayed by ED38 and ei-JB40 may not be specific binding to multiple antigens. This assertion needed to be investigated further and to avoid confusing terminology we referred to this non-specific phenotype as “promiscuous”.

The promiscuous phenotype was studied further with the “GF antibodies” which showed a strong promiscuous phenotype when they were initially discovered (Figure 3-13, page 166), but for some of them the phenotype proved temperamental as they did not all show it again in further ELISAs (Figure 3-14, page 167). The GF antibodies showed no evidence of binding biotinylated insulin in a soluble ELISA format (Figure 3-16, page 168), indicating that despite the signal in the solid phase ELISA, these antibodies do not bind insulin (although it is possible that the binding site was blocked by the biotinylation). All but GF07 did not bind the HEp-2 nuclei in immunocytochemistry (Figure 3-17, page 169), indicating that the antibodies do not bind dsDNA either. However, all five GF antibodies showed strong cytoplasmic binding, indicating that they may be autoreactive.

These results were mirrored by the three antibodies selected from the B cell Development dataset which showed a strong promiscuous phenotype: CT0037, CT0053 and CT0054 (Figure 3-19, page 172). These three antibodies did not bind biotinylated insulin in the soluble ELISA (Figure 3-20, page 173) or HEp-2 nuclei in the immunocytochemistry (Figure 3-6 and Figure 3-8, pages 158 and 160 respectively). However, like the GF antibodies, all three promiscuous antibodies showed very strong HEp-2 cytoplasmic binding.

CT0037 had a CDR-H3 with a high aromatic amino acid content (section 3.3.4, page 150) and CT0053 and CT0054 had long, hydrophobic CDR-H3 (section 3.3.2, page 149). All of these properties have been linked with polyreactivity in the literature (Meffre et al., 2004, Ng et al., 2004, Berkowska et al., 2015). A very striking feature of these three highly promiscuous antibodies was their propensity to aggregate as shown by the HP-SEC trace generated during their purification (Figure 3-21, page 173). This suggests that, alongside cytoplasmic binding, the promiscuous phenotype may be closely linked to antibody aggregation. All three antibodies also had unusual thermal shift assay traces (Figure 3-22, page 174); this may have been due to the disintegration of aggregates.

The promiscuous phenotype described in this chapter has been observed in human serum. Some sera being tested for diagnostic purposes do display background interactions in wells lacking test antigen (Fujii et al., 1989, Jørgensen et al., 2005, Terato et al., 2014, Güven et al., 2014) and sometimes appear to bind multiple autoantigens when coated on an ELISA plate, but not in other formats (Güven et al., 2014). Further, changing the blocking conditions can alter this background binding. For example, certain protein blocks (Terato et al., 2014) or a high-salt blocking buffer (Jørgensen et al., 2005) can remove this phenotype from serum as it did with ED38 and ei-JB40 in Figure 3-12 (page 165). The promiscuous phenotype in serum correlates with inflammatory mediators and immune complexes and could be induced by heating the serum to 40^{o}C (Güven et al., 2014). However, we observed this phenotype in purified antibodies, suggesting that other serum components are not required, but we did notice an association with aggregation propensity (Figure 3-21). Although the visibility of the promiscuous phenotype on the solid-phase ELISA is dependent on the blocking buffer used, the fact that not every antibody shows this phenotype (for example m-GO53) indicates that the phenotype is not simply caused by antibody protein sticking to a poorly blocked ELISA well.

In an attempt to isolate polyreactive antibodies from the human repertoire which do bind specifically to multiple unrelated antigens, we used phage display. The BMVtrp human scFv library was used in a panning selection against four unrelated autoantigens (HEp-2 cell lysate, dsDNA, MPO and ApoH). To prevent the isolation of promiscuous scFv, the selections were blocked using 3% (w/v) milk powder in PBS. Multiple selection cascades were used to mitigate the risk of one failing due to the order of the antigens (Figure 3-24, page 177). Unfortunately, no polyreactive antibodies were isolated from any of the selection cascades. We do not think that there was a problem with the phage display protocol as we did manage to isolate one anti-dsDNA antibody from the selection, however, we do acknowledge that the selection may not have been optimal; we may have found some polyreactive scFv if more colonies had been picked per round, different antigens had been used or a different phage display library had

been used. However, it is worth noting that the BMVtrp library contained sequences from the human bone marrow (Vaughan et al., 1996). If 55.2% of immature B cells are polyreactive (Wardemann et al., 2003), one would expect some polyreactive scFv to be isolated from the BMVtrp library using this selection cascade.

Despite these findings, we are not advocating that polyreactive antibodies do not exist. Polyreactivity can be caused by molecular mimicry (Oldstone, 1998) and it has been shown with X-ray crystallography that some antibodies can bind to multiple unrelated antigens by making different specific contacts (James et al., 2003, Kramer et al., 1997, Fields et al., 1995). However, these antibodies do not bind every antigen against which they are tested and it is unknown how common they are in the human repertoire (although it is possible that all antibodies are polyreactive and the limiting factor for identifying them is the size of the panel of antigens against which they are tested).

From the results of this chapter, we are suggesting that polyreactive antibodies which can specifically bind to antigens as diverse as DNA, LPS and insulin are much rarer than was suggested in Wardemann et al. (2003). Instead, we propose that the phenotype observed in papers using the protocol set out in Tiller et al. (2008) or similar may in fact be promiscuity. We do not believe promiscuity to be specificity to multiple antigens or “stickiness” as we do not see evidence of antigen-binding outside of the solid-phase ELISA format.

The cause of the promiscuous phenotype needs to be investigated further to determine why it is physiologically important and what role it plays in the immune response. It appears that the promiscuous phenotype is heavily selected at multiple tolerance checkpoints and is associated with various autoimmune diseases. It also appears to be linked to autoreactivity as the promiscuous antibodies we investigated showed strong cytoplasmic binding.

Chapter 4

Light Chain Selection and Isotype Properties

Chapter 4: Light Chain Selection and Isotype Properties

4.1 Introduction

The antigen binding site of an antibody is composed of a heavy chain and a light chain. When investigating antibody specificity, the focus is often centred on the heavy chain, in particular the CDR-H3. However, the light chain is certainly important for antibody specificity and should not be overlooked. On average, the CDR-L1, CDR-L2 and CDR-L3 together contribute approximately 32% of the energetically important residues during antigen binding (Kunik & Ofran, 2013). Furthermore, light chains are swapped during receptor editing for the purpose of altering BCR specificity to rescue autoreactive B cells (Yachimovich et al., 2002, Halverson et al., 2004). We showed in Figure 3-5 (page 156) that antibodies with the same heavy chain but different light chains can alter the binding specificity of the antibody.

Like heavy chains, light chains are implicated in some autoimmune diseases (Dörner et al., 1998, Steinsbø et al., 2014) and therefore it follows that they may be subject to similar selection pressures during B cell development as heavy chain variable region sequences. Although a previous study has not found evidence of selection pressures on light chain properties during B cell development (Wardemann et al., 2003), this was based on a relatively small “low throughput” dataset.

Following on from the heavy chain data analysis presented in the previous chapter (section 3.2), we hypothesised that examining bone marrow and peripheral blood light chain variable region properties using a large high-throughput sequencing dataset may highlight properties under selective pressures not visible when examining smaller datasets. Therefore, we used the light chain B cell Development dataset to determine whether any light chain properties significantly increased or decreased in prevalence as B cells matured, indicating that they may be under selection from tolerance mechanisms. We also investigated whether the light chain properties of transitional B cells show the same non-linear pattern of B cell progression as was observed in the heavy chain B cell Development dataset.

Furthermore, human light chains can be one of two classes: kappa or lambda. Unlike heavy chain classes (IgA, IgD, IgE, IgG, IgM) which are swapped onto the same variable region during class switch recombination (section 1.7, page 42), kappa and lambda light chains are encoded by separate sets of variable and constant region genes encoded on separate chromosomes. Kappa light chains are encoded on chromosome 2 (Malcolm et al., 1982) and lambda light chains are encoded on chromosome 22 (Erikson et al., 1981). These two classes are not used

equally in the B cell repertoire. As kappa and lambda variable regions are encoded by separate sets of genes, this suggests that they may have different properties which may lead to different roles in the antibody response. Light chains are more rarely the focus of publications than heavy chains, however there is some evidence that kappa and lambda light chains have differing roles. During B cell development, the kappa locus is usually rearranged before the lambda locus (Hieter et al., 1981) and it is thought that lambda light chains are better at rescuing autoreactive heavy chains than kappa light chains (Wardemann et al., 2004). In the peripheral blood, there is a bias towards kappa light chains with approximately 3/5 to 2/3 of the B cell repertoire using kappa light chains as opposed to lambda (Mole et al., 1994, Brauninger et al., 2001, Montano & Morrison, 2002). However, in antigen-selected populations this ratio can differ (Chui et al., 1990) for example, lambda light chains are more prevalent than kappa in secreted (predominantly IgA) antibodies (Mole et al., 1994). Additionally, HIV Env-specific antibodies isolated from chronic HIV patients have a very strong bias towards lambda light chains (Sajadi et al., 2016) and hairy cell leukaemia patients also have a bias towards lambda light chains (Forconi et al., 2008). Phenotypic differences such as conformational flexibility (Stanfield et al., 2006) and half-life (Montano & Morrison, 2002) have also been found between antibodies bearing kappa and lambda light chains.

These differences in kappa and lambda frequency and functionality and the fact that they are encoded by completely separate sets of variable and constant region genes suggests that kappa and lambda light chains may have different roles in the immune system. We hypothesised that if kappa and lambda light chains do have different roles in the immune response (e.g. binding different types of antigen), this may be reflected in their physicochemical and/or structural properties. Therefore, we analysed the CDRs of kappa and lambda light chains to determine whether there were physicochemical and/or structural differences in the variable regions of the two classes.

Furthermore, there is much debate around whether immunoglobulin heavy-light chain pairing is entirely random or whether there are biases in which heavy and light chains pair together. We hypothesised that if we saw differences in the physicochemical/structural properties of kappa and lambda light chains, this may be reflected by physicochemical/structural differences in the heavy chains with which kappa and lambda light chains are paired. If this were the case it would suggest that heavy-light chain pairing is not random.

Finally, we investigated whether the mean amount of random N-nucleotide addition in the CDR3 regions differed between individuals, and whether there was any correlation between the amount of N-nucleotide addition in the heavy and light chain CDR3 regions.

In summary, the aim of this chapter is to further our understanding of the properties of immunoglobulin light chains and heavy-light chain pairing using the following methods:

1. Analysis of the light chain B cell Development dataset to identify variable region properties which increase or decrease in frequency as B cells mature as these properties may be associated with autoreactivity
2. Compare the physicochemical properties of kappa and lambda CDR-L3 in the light chain B cell Development dataset to determine if they are significantly different, suggesting that kappa and lambda light chains may bind better to different types of antigen
3. Use a published, paired heavy-light chain high-throughput sequencing dataset to compare the physicochemical properties of CDR-H3 of heavy chains paired with kappa and lambda light chains to determine if there is evidence of non-random heavy-light chain pairing
4. Investigate the structural properties of all the CDRs of kappa and lambda antibodies in the Protein Data Bank (PDB) to determine if there are structural differences between kappa and lambda CDRs or the CDRs of the heavy chains with which they are paired
5. Investigate the mean N-nucleotide addition at the kappa, lambda and heavy chain CDR3 regions in each donor in the B cell Development dataset to determine if there is any correlation in the mean number of N-nucleotides added

4.2 Light Chain Properties under Selection as B cells Mature

4.2.1 VJ Gene Usage in Bone Marrow and Peripheral Blood B cell Repertoires

The mean frequency of usage of each light chain family/gene (+/- 1 SD) in the immature, transitional and naïve B cell repertoires was calculated for each of the 19 donors in the light chain B cell Development dataset. Significant differences in family/gene frequency between the B cell repertoires were calculated using two-way unpaired ANOVA followed by Tukey's multiple comparisons post-test.

The analysis was performed on the entire light chain dataset which contained 30,161 kappa and 14,183 lambda entries (Table A-21, page 275), however a small number of samples were excluded due to there being too few sequences to accurately calculate family/gene usage (see section 2.5.1, page 96, for details of the "5X rule"). The samples excluded from analysis are shown in Table 4-1.

Table 4-1: Samples excluded from VJ family usage analysis due to "5X rule". Samples excluded from V family usage analysis are shown in red (<35 kappa entries/ <50 lambda entries). Donors excluded from J family usage are underlined (<25 kappa entries/ <35 lambda entries). A red dash indicates samples for which no data was available.

	Kappa			Lambda		
Donor ID	Immature	Transitional	Naïve	Immature	Transitional	Naïve
122	1380	349	470	308	52	87
159	1292	419	532	349	140	135
138	255	130	961	22	125	-
107	-	1476	163	-	686	152
120	-	841	378	-	568	212
118	464	249	281	348	88	58
146	948	485	131	516	109	37
103	371	15	554	164	3	307
128	567	360	-	400	234	-
126	259	664	1651	83	242	897
141	2021	475	478	1414	233	73
105	288	183	417	170	83	303
119	2175	666	334	1609	27	162
160	304	130	757	207	71	37
111	188	151	708	79	67	644
149	848	-	246	689	-	171
132	-	896	515	-	381	267
162	-	1139	1	-	1	249
140	553	279	764	368	174	382

Very little difference was seen in kappa or lambda VJ family/gene usage between the immature, transitional and naïve B cell repertoires. Like the heavy chain analysis, there were several examples of a gene being used significantly more or less frequently in the transitional repertoire compared to the other repertoires, supporting the hypothesis that B cells do not progress entirely linearly and that transitional B cells may become a variety of different cell types, subject to different selection pressures.

The only IGKV family which significantly differed in frequency between the cell repertoires was IGKV4 which was slightly less frequent in the naïve repertoire compared to the immature repertoire (Figure 4-1A). This may have been driven by the IGKV4-1 gene, which became significantly less frequent as B cells matured. IGKV1-39 and IGKV3-11 became significantly more frequent as B cells matured and IGKV3-20 was significantly less frequent in the transitional and naïve repertoires compared to the immature repertoire. IGKV1-33 was significantly less frequent in the transitional repertoire compared to the immature and naïve repertoires.

Lambda light chains had slightly more significant differences between cell repertoires than kappa. IGLV1 became significantly less frequent and IGLV2 became significantly more frequent as B cells matured (Figure 4-1B). IGLV2-23 was significantly more common in the transitional and naïve repertoires compared to the immature repertoire and IGLV2-14 was significantly more common in the naïve repertoire compared to the immature repertoire, but the usage in the transitional repertoire was significantly lower than either the immature or naïve repertoires. IGLJ1 was significantly more common and IGLJ3 was significantly less common in the naïve repertoire compared to the immature repertoire.

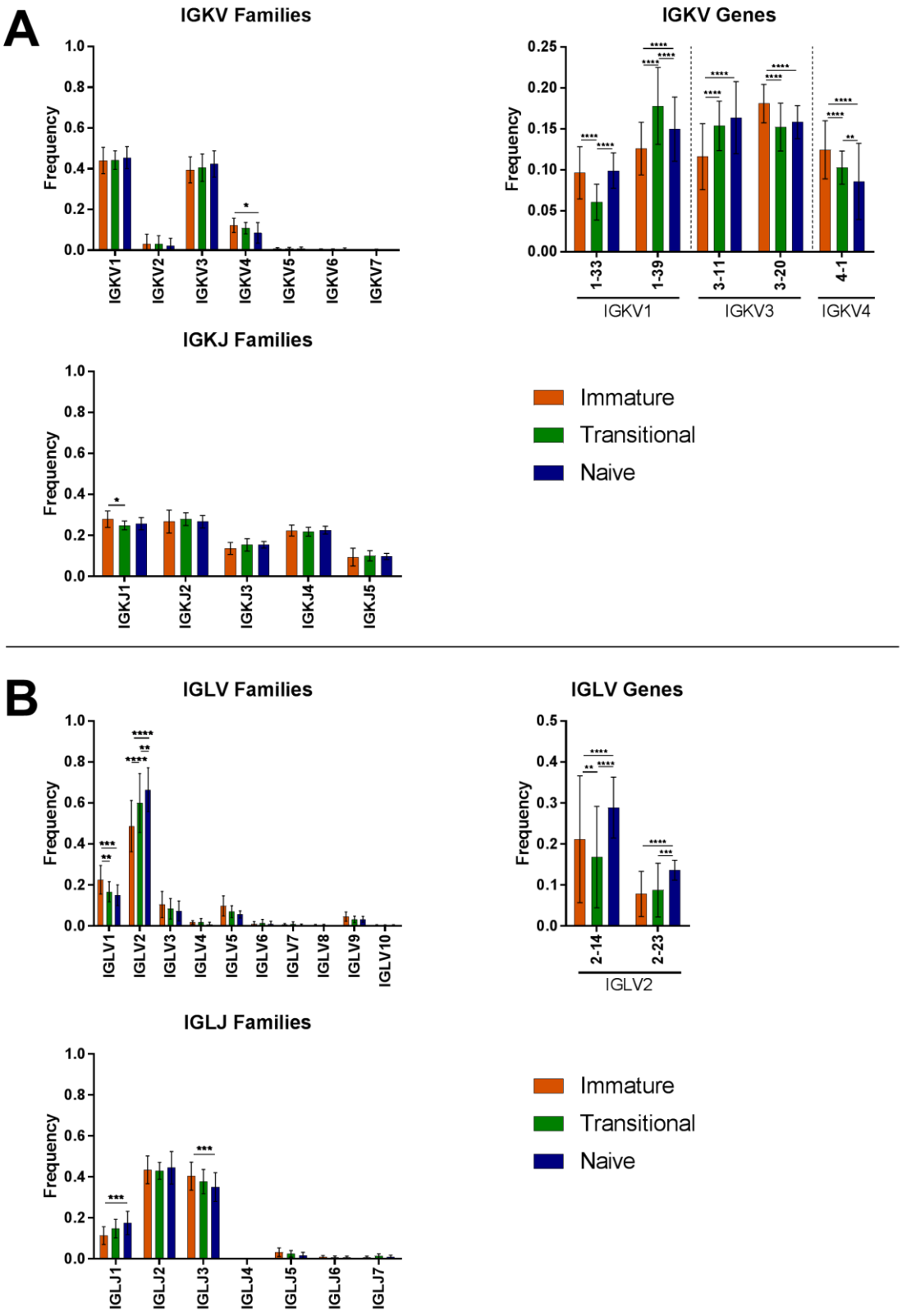


Figure 4-1: Light chain VJ family and gene frequencies in each cell repertoire. Mean frequency (+/- 1 SD) of (**A**) kappa and (**B**) lambda families/genes in immature, transitional and naïve B cell repertoires. Only V genes which differed significantly between cell types are shown. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

4.2.2 CDR-L3 Properties in Bone Marrow and Peripheral Blood B cell Repertoires

Before performing analysis of the CDR-L3 regions, sequence were removed from the light chain B cell Development dataset which may skew the analysis (section 2.5.2, page 96). The resulting dataset contained 20,571 kappa and 8,876 lambda entries (Table A-22, page 275).

The following physicochemical properties were calculated (section 2.4.2, page 94) for each CDR-L3 amino acid sequence:

1. CDR-L3 length (amino acids)
2. Aliphatic index
3. Boman index
4. Isoelectric point
5. Hydrophobicity (GRAVY index)
6. Instability index

Data for the 19 donors was pooled and frequency distributions and the mean (+/- 1 SD) were calculated for each cell type (immature, transitional and naïve). The Kruskal-Wallis test followed by Dunn's multiple comparisons post-test was used to identify significant differences between the cell types.

There were very few significant differences in the kappa CDR-L3 properties between the three cell types. The only significant difference was in the mean isoelectric point (pI). Like the heavy chain repertoire (Figure 3-2D, page 142), the mean pI of the transitional repertoire was significantly higher than the mean pI of the immature and naïve repertoires (Figure 4-2D).

There were more significant differences in the CDR-L3 properties of the lambda repertoire than the kappa repertoire. The aliphatic index was significantly lower in the naïve repertoire compared to the immature and transitional repertoires (Figure 4-3B). The naïve repertoire also had more hydrophobic (higher GRAVY index) CDR-L3 regions on average than the immature and transitional repertoires (Figure 4-3E). The Boman index was significantly higher in the transitional repertoire than the immature repertoire, but was not different to the naïve repertoire (Figure 4-3C). The mean CDR-L3 length of the transitional repertoire was significantly shorter than the immature repertoire, bit not different to the naïve repertoire (Figure 4-3A). The mean pI was significantly higher in the naïve repertoire than the immature repertoire (Figure 4-3D). The CDR-L3 instability index was significantly higher in the transitional and naïve repertoires compared to the immature repertoire (Figure 4-3F).

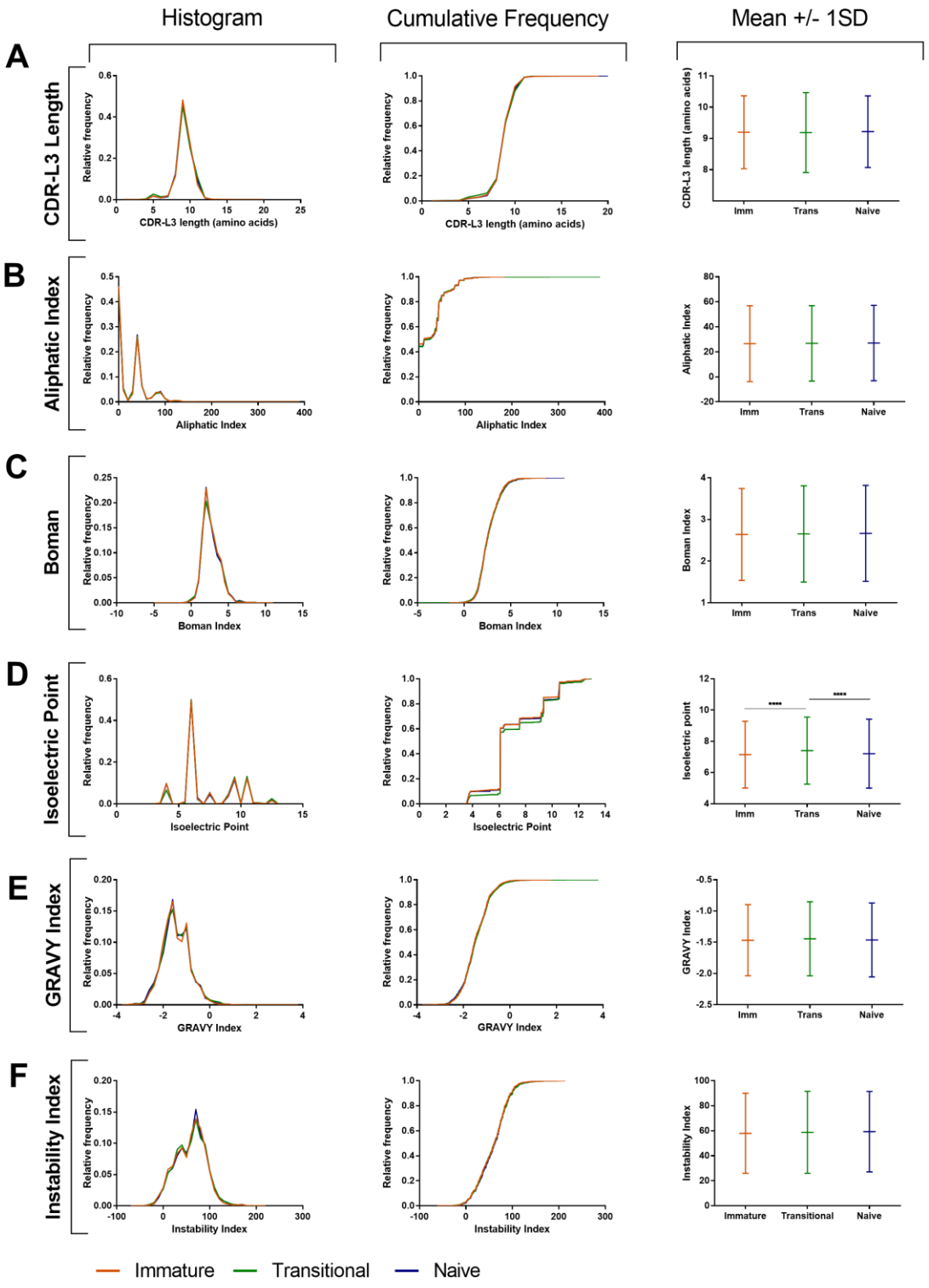


Figure 4-2: Kappa CDR-L3 physicochemical properties in each cell type. Histogram, cumulative frequency histogram and mean (+/- 1 SD) for six physicochemical properties of the kappa CDR-L3 regions of B cells isolated from the immature, transitional and naïve repertoires. (**A**) CDR-L3 length; (**B**) Aliphatic index; (**C**) Boman index; (**D**) Isoelectric point; (**E**) GRAVY index; (**F**) Instability index. Each set of three graphs shows the same data in different formats. The mean +/- 1 SD graphs were used to see significant differences between cell types, whereas the histograms and cumulative frequency plots show the shape of the data. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

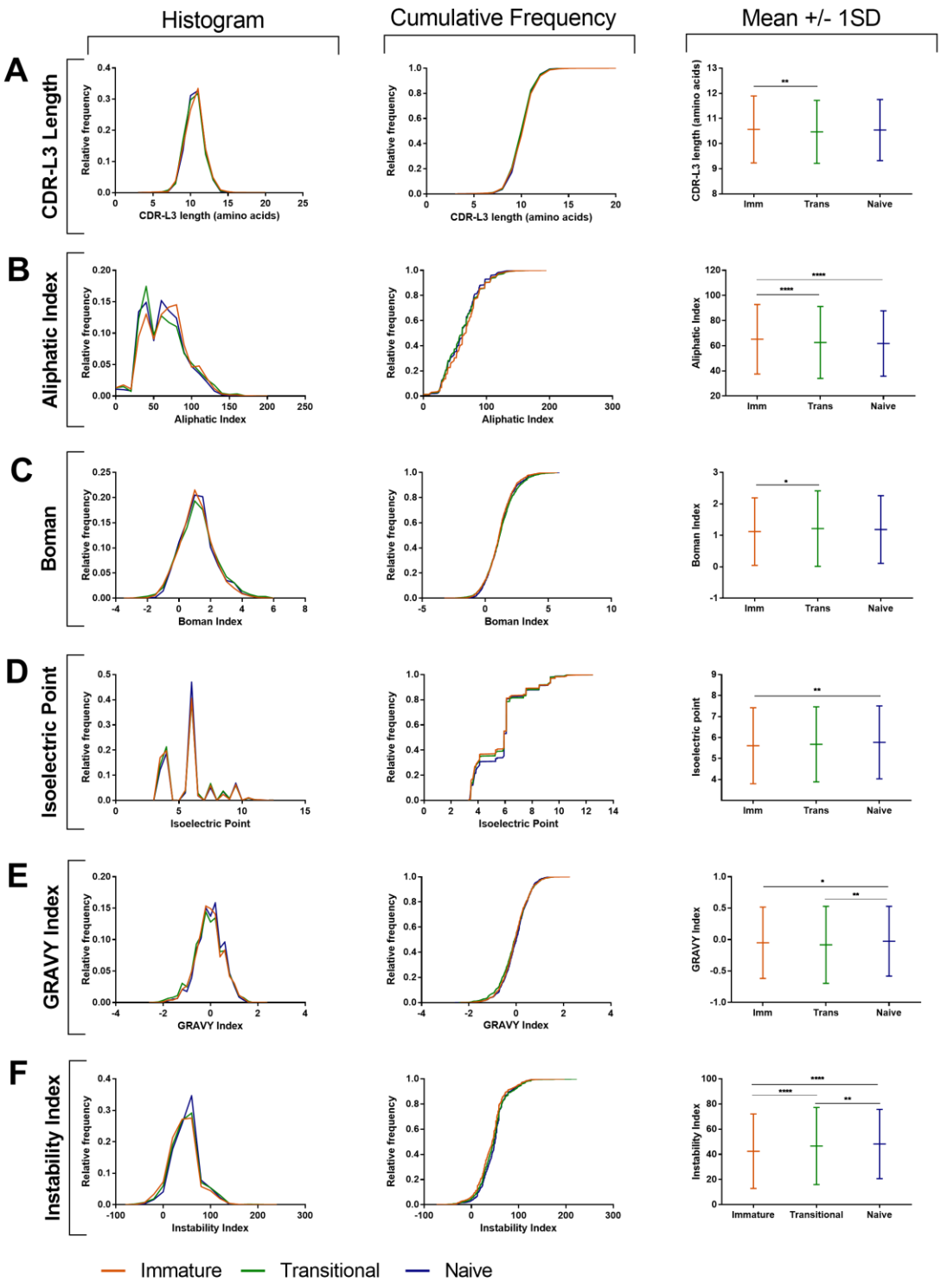


Figure 4-3: Lambda CDR-L3 physicochemical properties in each cell type. Histogram, cumulative frequency histogram and mean (+/- 1 SD) for six physicochemical properties of the lambda CDR-L3 regions of B cells isolated from the immature, transitional and naïve repertoires. (**A**) CDR-L3 length; (**B**) Aliphatic index; (**C**) Boman index; (**D**) Isoelectric point; (**E**) GRAVY index; (**F**) Instability index. Each set of three graphs shows the same data in different formats. The mean +/- 1 SD graphs were used to see significant differences between cell types, whereas the histograms and cumulative frequency plots show the shape of the data. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

4.3 Kappa and Lambda CDRs Have Different Physicochemical Properties

In section 4.2.2 (page 195), the light chain CDR-L3 physicochemical properties were compared between the immature, transitional and naïve cell repertoires to determine which properties may be under selection as B cells matured. In this section, the kappa and lambda CDR physicochemical properties were compared with each other to investigate whether these two light chain classes have significantly different CDR properties.

Firstly, the physicochemical properties of kappa and lambda CDR-L3 regions in each cell repertoire (immature, transitional and naïve) were compared. The same light chain dataset was used as in section 4.2.2 (page 195) and the kappa and lambda CDR-L3 physicochemical properties were plotted as cumulative frequency histograms. The Kolmogorov-Smirnov (KS) test was used to test for significant differences between each kappa and lambda repertoire.

In all three cell repertoires the kappa and lambda CDR-L3 had very different physicochemical properties (Figure 4-4A). The lambda CDR-L3 were significantly longer, more hydrophobic and had a higher aliphatic index than the kappa CDR-L3. The kappa CDR-L3 had higher Boman indexes and higher isoelectric points (pI). Interestingly, like the heavy chain pI cumulative frequency graphs (Figure 3-2D, page 142), both kappa and lambda repertoires had very few CDR-L3 with a pI in the range either side of approximately 6.0. However, the lambda repertoire had a higher frequency CDR-L3 with a low pI and the kappa repertoire had a higher frequency of CDR-L3 with a high pI.

The frequency of each type of amino acid (Table 2-20, page 95) in kappa and lambda CDR-L3 was investigated. The mean frequency (+/- 1 SD) of each type of amino acid was plotted and significant differences between the kappa and lambda repertoires were found using multiple *t*-tests followed by the FDR correction for multiple comparisons. The kappa and lambda repertoires had very different CDR-L3 amino acid usage (Figure 4-4B). The lambda repertoire had a significantly higher frequency of small, tiny, non-polar, aliphatic, charged and acidic amino acids, whereas the kappa repertoire has a significantly higher frequency of polar, aromatic and basic amino acids.

The 10 Kidera factors (Table 2-19, page 94) were calculated for each CDR-L3 and these 10 orthogonal measures of physicochemical and structural properties were used in a principle component analysis (PCA) (Figure 4-4C). The PCA separated the kappa and lambda repertoires into distinct groups. Kidera 10, Kidera 2, Kidera 5 and Kidera 6 were the factors which separated the kappa and lambda classes most.

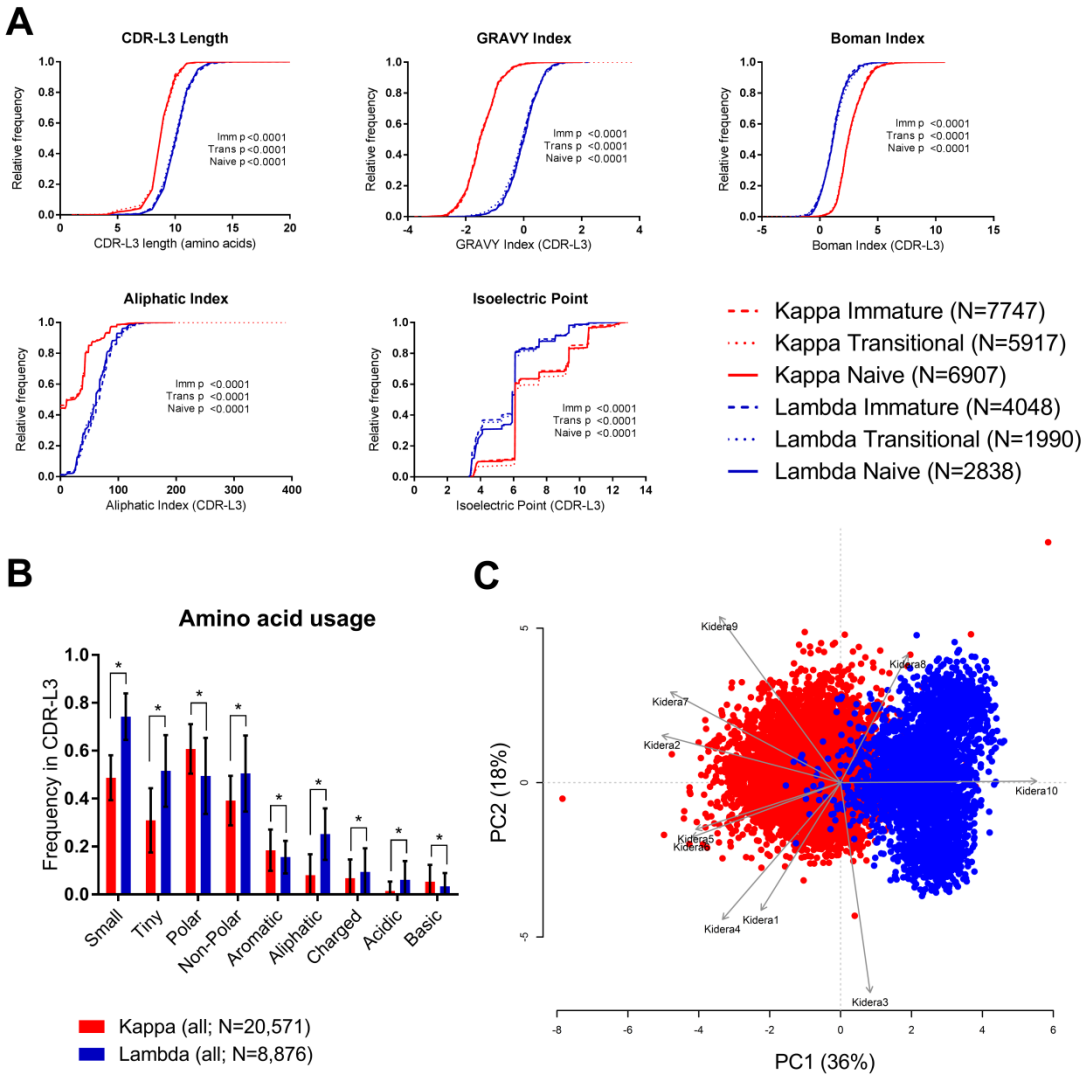


Figure 4-4: Physicochemical properties of kappa and lambda CDR-L3. All graphs show the CDR-L3 properties of kappa (red) and lambda (blue) CDR-L3 of B cells isolated from the immature, transitional and naïve repertoires of 19 healthy donors. (**A**) The physicochemical properties of kappa and lambda CDR-L3 plotted as cumulative frequency graphs. (**B**) Mean frequency (+/- 1 SD) of each group of amino acids (see Table 2-20, page 95) in kappa and lambda CDR-L3 (asterisk indicates significance). (**C**) PCA using the ten Kidera factors (Table 2-19, page Table 2-1994) calculated for kappa and lambda CDR-L3.[94]

[94] The graphs shown in Figure 4-4 have been published in Townsend et al. (2016). The R script used to perform the PCR shown in Figure 4-4C was written by Grace Lu (KCL).

CDR-L3 are encoded by germline V and J genes and also random N-nucleotide addition. This raises the question of whether the physicochemical differences seen in the complete CDR-L3 (Figure 4-4) are due to differences encoded by the V and J gene segments, or by the N-nucleotide addition.

To investigate this, a dataset of "germline" CDR-L3 which lacked N-nucleotide addition/deletion was created by combining germline light chain V and J gene amino acid sequences in the same frequency with which each V-J pair was seen in the B cell Development dataset. To do this, IMGT Protein Displays (Scaviner et al., 1999) were used to obtain the amino acid sequences of the CDR-L3 regions of germline V genes and J genes (i.e. V gene position 105 onwards and the first two amino acids of the J gene). Only *01 alleles were used.

The frequency of recombination of each V gene and J gene in the real datasets (20,571 kappa and 8,876 lambda entries) was determined. The CDR-L3 sections of the V gene and J gene germline sequences were combined at the equivalent frequency. 428 entries were removed due to the resulting CDR-L3 regions being unproductive. The final dataset contained 20,379 kappa and 8,648 lambda entries.

The physicochemical properties of each "germline" CDR-L3 was calculated using the R package Peptides as described in section 2.4.2 (page 94). The kappa and lambda CDR-L3 properties were then compared as described for Figure 4-4A.

The same differences in physicochemical properties were seen between the kappa and lambda "germline" CDR-L3 as were seen in the real-life dataset (Figure 4-5). This indicates that the differences in CDR-L3 physicochemical properties seen between kappa and lambda are encoded in the germline and are not the result of differences in N-nucleotide addition. However, the physicochemical properties of the "germline" CDR-L3 were significantly different from their real life counterparts indicating that N-nucleotide addition does significantly influence the physicochemical properties of CDR-L3, however the lack of directionality in this difference indicates that it does so randomly.

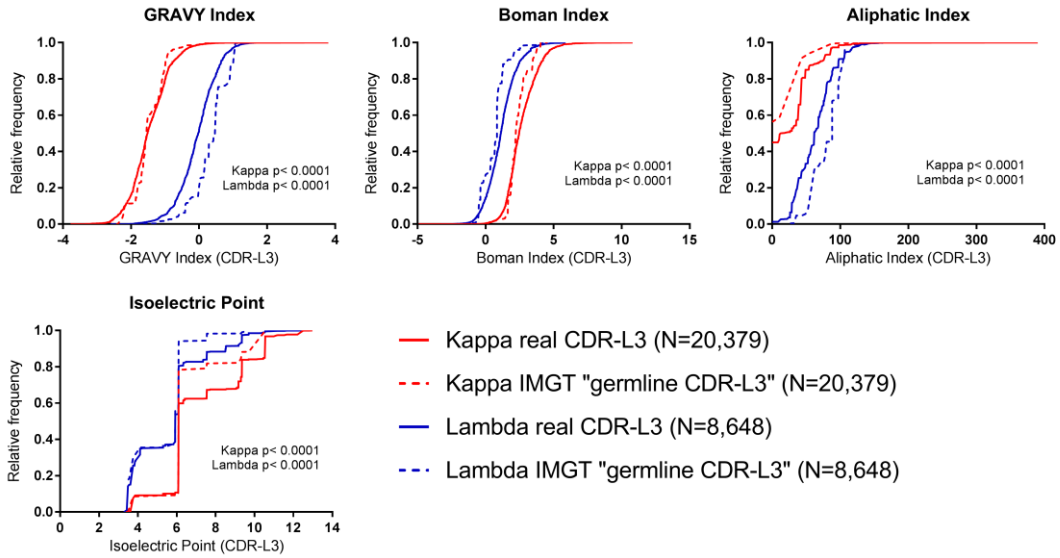


Figure 4-5: Physicochemical properties of kappa and lambda "germline" CDR-L3. CDR-L3 physicochemical properties of real (solid line) and germline (dotted line) kappa (red) and lambda (blue) CDR-L3 plotted as cumulative frequency graphs. Real CDR-L3 data from the immature, transitional and naïve repertoires of 19 healthy donors was pooled for this figure.[95]

[95] The graphs shown in Figure 4-5 have been published in Townsend et al. (2016).

Cumulative frequency graphs and the KS test were used to determine if there were any significant differences in the physicochemical properties of CDR-L1 or CDR-L2 (IMGT definition) between the kappa and lambda repertoires. Significant differences were observed (Figure 4-6), however the differences are smaller than those seen in the CDR-L3 regions and, due to lower diversity in the CDR-L1 and CDR-L2 regions, the graphs are not as smooth.

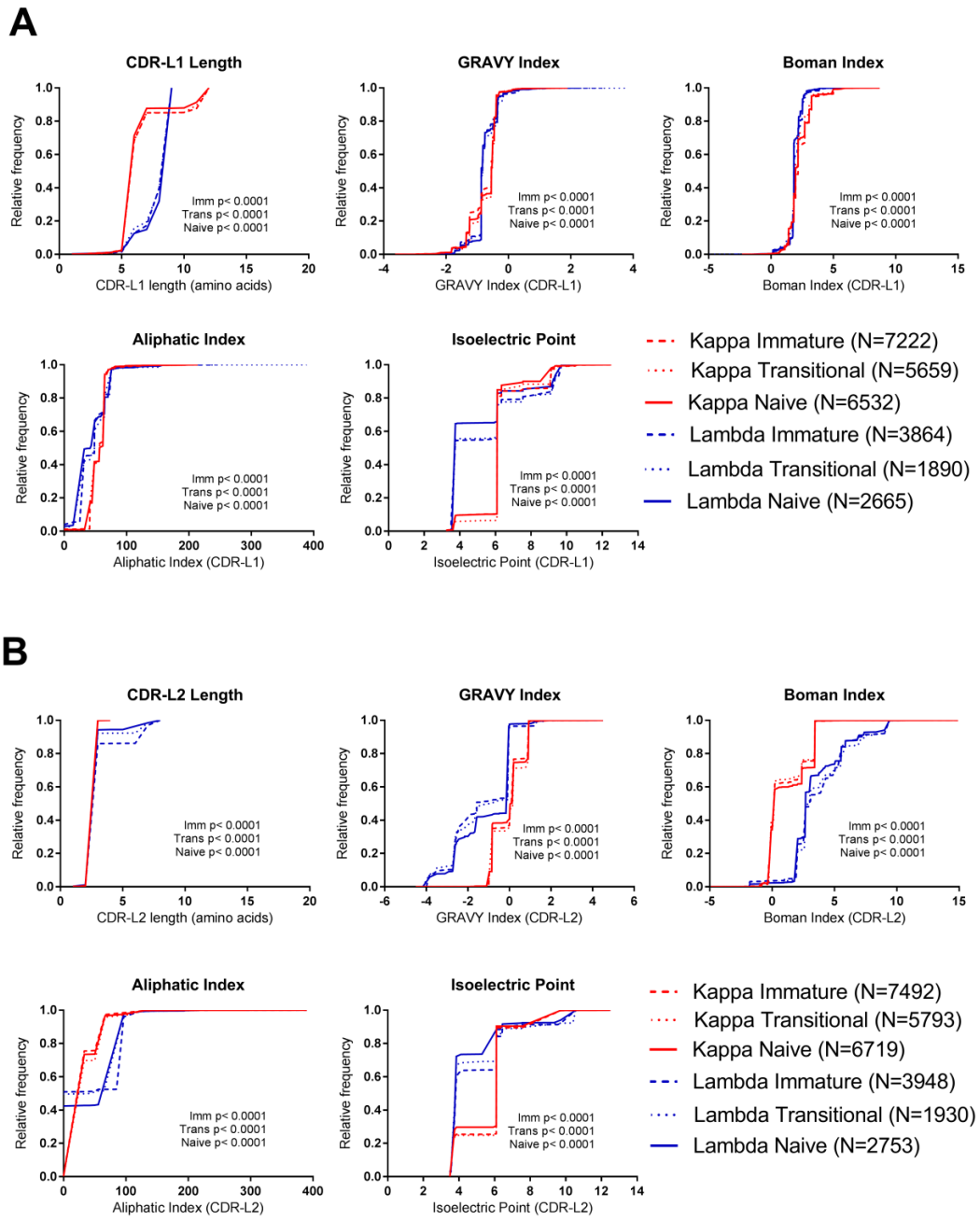


Figure 4-6: Physicochemical properties of kappa and lambda CDR-L1 and CDR-L2. The physicochemical properties of kappa (red) and lambda (blue) (**A**) CDR-L1 and (**B**) CDR-L2 regions isolated from immature, transitional and naïve B cells of 19 healthy donors plotted as cumulative frequency graphs.[96]

[96] The graphs shown in Figure 4-6 have been published in the supplemental material of Townsend et al. (2016).

4.4 CDR-H3 of Kappa and Lambda Antibodies are not Different

In order to compare the CDR-H3 physicochemical properties of heavy chains paired with kappa and lambda light chains, heavy-light chain pairing information was needed. The B cell Development dataset does not contain heavy-light chain pairing information, so a published dataset was used (DeKosky et al., 2016). This dataset was generated by the George Georgiou lab using their technique of paired heavy-light chain sequencing (DeKosky et al., 2015, McDaniel et al., 2016) and comprises paired heavy and light chain sequences from the antigen-naïve repertoires of three healthy donors. Donor 1 had 13,771 sequences, Donor 2 had 26,343 sequences and Donor 3 had 15,193 sequences. We split the heavy chain dataset based on whether the heavy chains were paired with kappa or lambda light chains.

Physicochemical properties for each CDR-H3 were calculated as described in section 2.4.2 (page 94) and the CDR-H3 physicochemical properties of heavy chains paired with kappa and lambda light chains were compared using cumulative frequency graphs and the KS test.

VDJ family usage in the kappa- and lambda-paired heavy chains was investigated by plotting the mean frequency (+/- 1 SD) of each family across the three donors. Significant differences between kappa- and lambda-paired heavy chain repertoires were tested using multiple *t*-tests followed by FDR correction for multiple testing.

A significant difference was seen in the CDR-H3 isoelectric points (pI) of Donor 1 and Donor 2, and in the CDR-H3 length of Donor 2. No other significant differences in CDR-H3 physicochemical properties were observed (Figure 4-7A). No significant differences in the mean VDJ family usage was observed between kappa and lambda repertoires (Figure 4-7B).

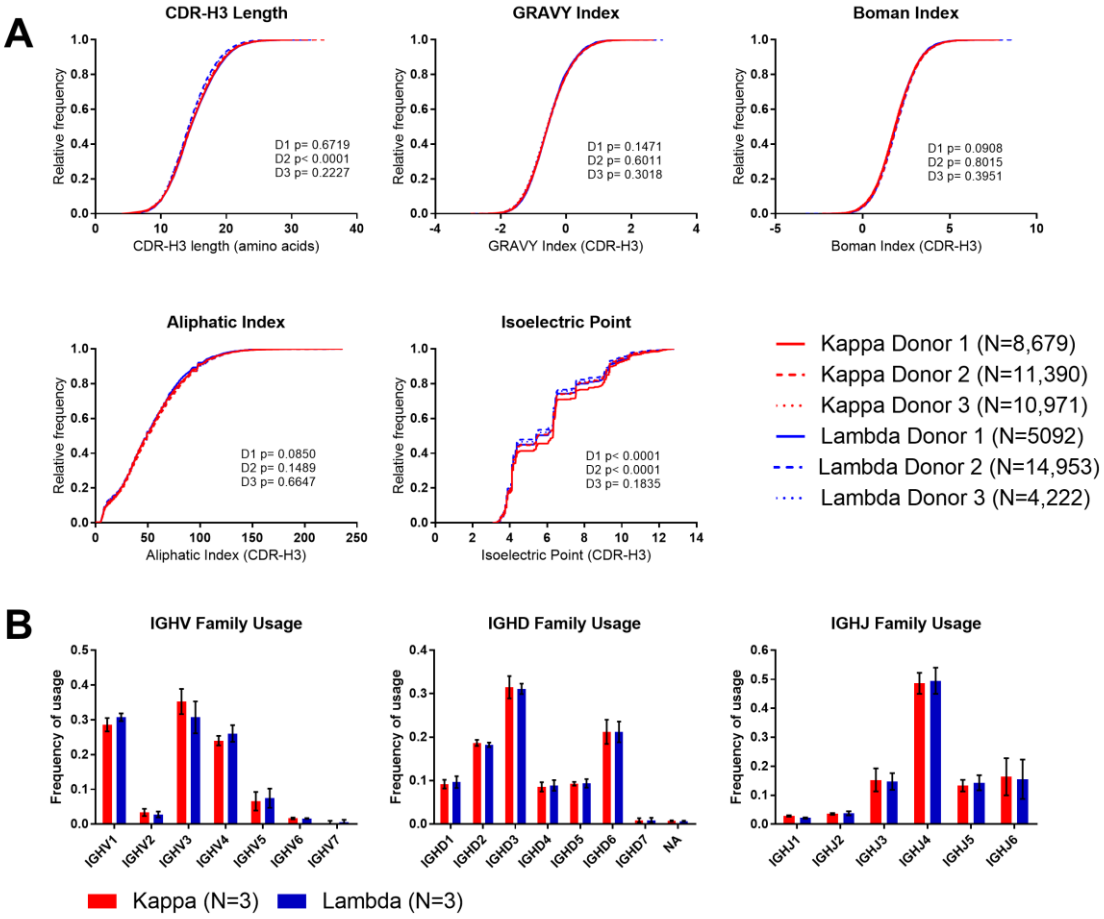


Figure 4-7: CDR-H3 properties of heavy chains paired with kappa or lambda light chains. All graphs show the CDR-H3 properties of heavy chains paired with kappa (red) and lambda (blue) light chains. Data was taken from the antigen-naïve dataset published in DeKosky et al. (2016). (**A**) The physicochemical properties of kappa and lambda CDR-H3 of each donor plotted as cumulative frequency graphs. (**B**) Mean frequency (+/- 1 SD) of IGHV, IGHD and IGHJ family usage in the three donors in heavy chains paired with kappa and lambda light chains.[97]

[97] The graphs shown in Figure 4-7 have been published in Townsend et al. (2016).

4.5 Kappa and Lambda CDR Loops are Structurally Different

In Figure 4-4C (page 199), the kappa and lambda repertoires could be separated by PCA based on their CDR-L3 Kidera factors, which are linked to structural properties. This suggests that in addition to physicochemical differences, there are structural differences in the CDR loops.

Antibody structures from the Protein Data Bank (PDB) were used to determine if there are structural differences between kappa and lambda classes in the CDR-L1, CDR-L2 and CDR-L3. For full materials and methods, see section 2.6, page 98. Briefly, all of the high-quality paired heavy-light chain crystal structures in the PDB were downloaded and redundancy was eliminated. 199 kappa and 106 lambda structures remained. Due to some incomplete CDR information, kappa antibody CDR-H1, lambda antibody CDR-H2, and lambda antibody CDR-L2 analyses were performed using 197, 105, and 103 entries respectively. The secondary structure occupancies of each heavy and light chain CDR (Chothia definition) were calculated for each of the PDB structures included in the analysis. The mean secondary structure occupancies (+/- 95% CI) were calculated for each CDR. Differences in the mean probability of secondary structures between kappa and lambda antibodies were considered significant if the 95% confidence intervals did not overlap.

The secondary structures investigated were Beta, Coil, Helix and Turn (see Table 2-21, page 98, for definitions). Example CDR-L3 loops with high Beta, Coil and Helix content are shown in Figure 4-8. Turn structure was not considered meaningful as this structure serves to link more regular secondary structure types.

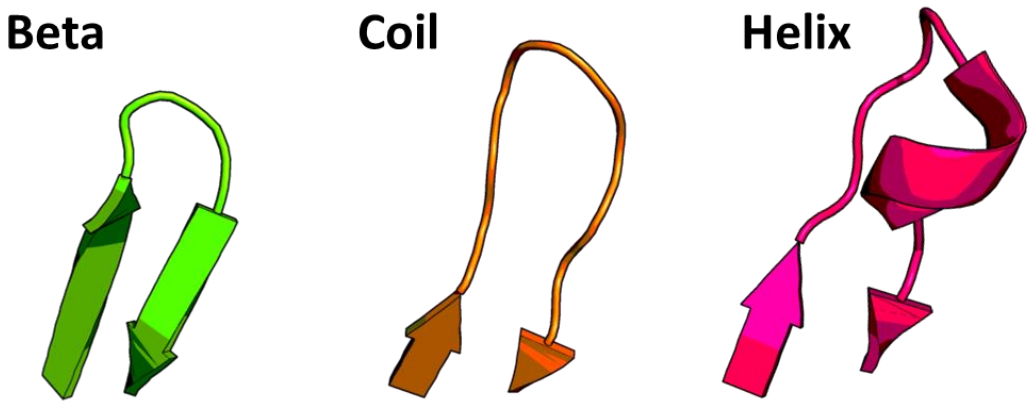


Figure 4-8: Example CDR-L3 loops with high Beta, Coli and Helix structures. See Table 2-21 (page 98) for definitions of Beta, Coil and Helix. CDR-L3 loops are from the following PDB structures: 4UT7 (Beta), 4ZD3 (Coil) and 5DT1 (Helix).

The Beta content of kappa CDR-L1 and CDR-L2 was significantly higher than lambda, but the Beta content of kappa CDR-L3 was significantly lower than lambda. In the CDR-L2 and CDR-L3, these differences in Beta content were compensated for by differences in Coil content, whereas in the CDR-L1, the differences in Beta content were compensated for by differences in Helix content (Figure 4-9A).

The only significant structural difference observed in the heavy chains paired with kappa or lambda light chains was a significantly lower probability of Helix content in the CDR-H2 of heavy chains paired with lambda light chains (Figure 4-9B). No other significant structural differences were found in CDR-H1, CDR-H2 or CDR-H3.

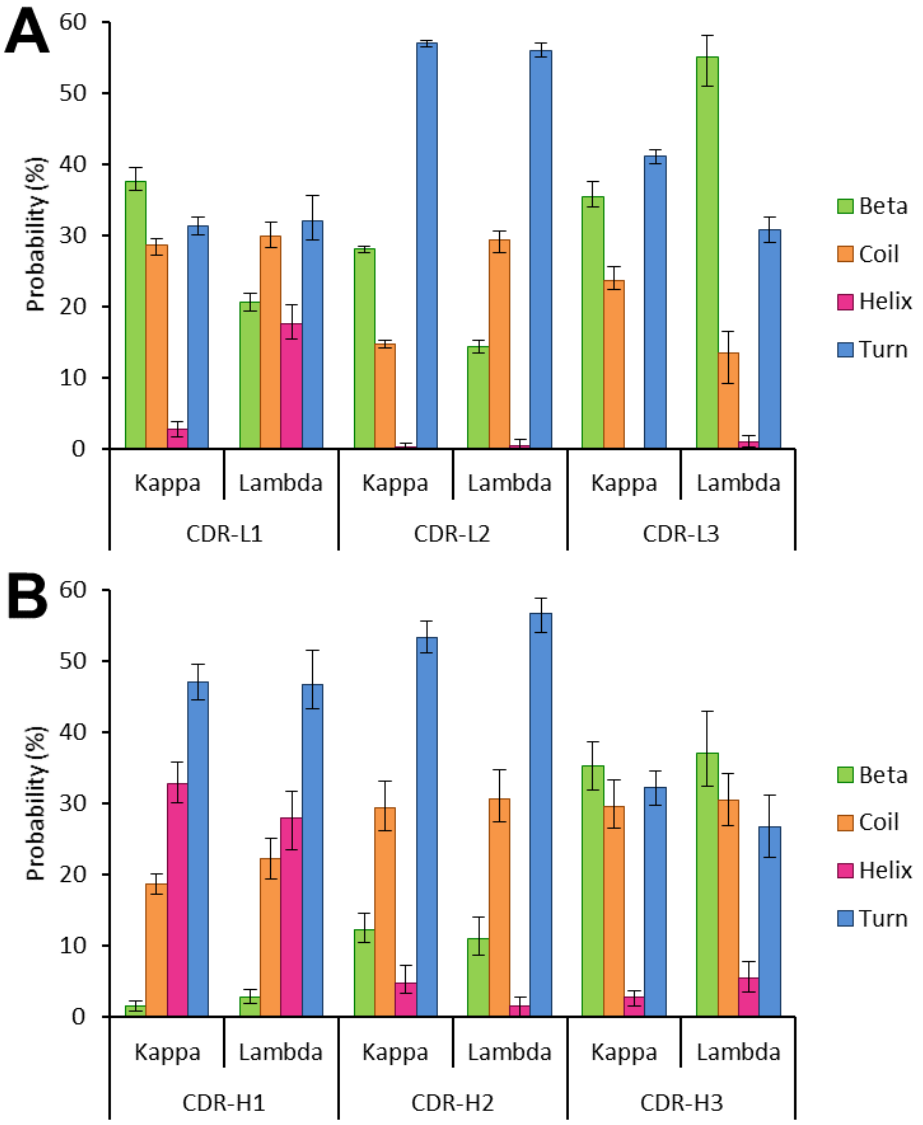


Figure 4-9: Structural differences between kappa and lambda CDR loops. The mean probability of Beta, Coil, Helix and Turn secondary structures (+/- 95% CI) occurring in the six CDR loops was plotted for kappa and lambda antibody structures extracted from the Protein Data Bank.[98]

[98] The data plotted in Figure 4-9 was generated by Julie Laffy (KCL). The graphs shown in this figure have been published in Townsend et al. (2016).

4.6 CDR3 N-Nucleotide Addition is Positively Correlated within Individuals

This chapter has demonstrated that there are significant physicochemical and structural differences between kappa and lambda CDR-L3. Figure 4-5 showed that these differences are primarily encoded in the germline rather than by random N-nucleotide addition. The number of N-nucleotides in light chain CDR-L3 regions is too low to investigate the physicochemical properties of the amino acids they encode. However, we could investigate whether there was any difference in the number of N-nucleotides added at the kappa and lambda CDR-L3.

We found that the amount of N-nucleotide addition was on average slightly higher in lambda CDR-L3 compared to kappa, however the Wilcoxon matched-pairs signed rank test showed that the difference was not significant ($p > 0.999$).

However, when the mean N-nucleotide addition at the kappa and lambda CDR-L3 for each donor was plotted on an X-Y graph, there was a very clear and highly significant positive correlation (Figure 4-10A). Furthermore, when the mean N-nucleotide addition at the kappa and lambda CDR-L3 regions was plotted against the mean N-nucleotide addition at the CDR-H3 within the same donors, there was also a significant positive correlation (Figure 4-10B&C respectively). Spearman's rank correlation coefficient was used to investigate the correlation.

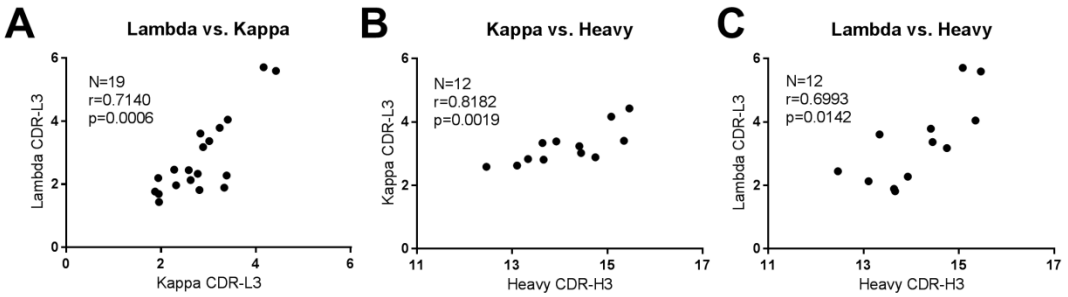


Figure 4-10: Positive correlation of mean number of N-nucleotides within individuals. Mean number of N-nucleotides in the kappa, lambda and heavy chain CDR3 regions of each donor are plotted. Number of donors indicated by "N". Spearman's rank correlation coefficient indicated by "r". Two-tailed p values indicated by "p".[99]

There was a very large range of mean N-nucleotide addition at each CDR3. The difference between the donors with the lowest and highest mean N-nucleotide addition at the kappa, lambda and heavy CDR3 regions was 2.5 nt, 4.3 nt and 3.0 nt respectively. There was no correlation between the amount of N-nucleotide addition and the age of the donor.

[99] The graphs shown in Figure 4-10 have been published in Townsend et al. (2016).

4.7 Discussion

4.7.1 Light Chain Variable Regions show little Evidence of Selection

As with the heavy chain analysis, we hypothesised that light chain variable region properties which decreased in frequency in the B cell Development dataset as the B cells matured may do so because they are selected out at tolerance checkpoints due to autoreactivity. We hypothesised the converse for properties which increased in frequency.

In Chapter 3, it was found that the heavy chain properties of the transitional B cell repertoire do not always fit the pattern that would be expected if B cells matured linearly from pre-B >> immature >> transitional >> naïve (discussed in section 3.7.1, page 178). In this chapter, we found that this trend was also present in the light chain B cell Development dataset. IGKV1-33 and IGLV2-14 occurred significantly less frequently in the transitional repertoire compared to the immature and naïve repertoires whereas IGKV1-39 occurred significantly more frequently in the transitional repertoire compared to the immature and naïve repertoires (Figure 4-1, page 194). The isoelectric point of the kappa transitional CDR-L3 repertoire was significantly higher than the immature and naïve repertoires (Figure 4-2D, page 196).

Unlike the heavy chain variable region properties investigated in Chapter 3 (section 3.2, page 138), there was very little evidence of positive or negative selection of light chain variable region properties as B cells matured from immature to naïve.

In the kappa repertoire, the only IGKV family which was significantly less frequent in the naïve repertoire compared to the immature repertoire was IGKV4. As IGKV4 only contains one gene (IGKV4-1), this was due to the decrease in frequency of IGKV4-1 in both the transitional and naïve repertoires compared to immature (Figure 4-1A, page 194). This suggests that IGKV4-1 may be selected against by tolerance mechanisms. IGKV4-1 is known to be over-represented in the repertoire of people with SLE (Dörner et al., 1998) and is associated with coeliac disease (Steinsbø et al., 2014). Based on the frequency of IGKV4-1 in DNA vs. cDNA it has been previously speculated that the gene may be selected against at tolerance (Hehle et al., 2015), but only now have the frequencies in the bone marrow and peripheral blood B cells been compared, allowing us to conclude this with more certainty. Alongside IGKV4-1, IGKV3-20 also decreased in frequency as B cells matured, whereas IGKV3-11 increased in frequency. It was surprising that long kappa CDR-L3 regions did not decrease in frequency as B cells matured (Figure 4-2A, page 196) as it has been shown that kappa CDR-L3 regions longer than 11 amino acids are associated with rheumatoid arthritis (Samuels et al., 2005), however the mean kappa CDR-L3 length in the B cell Development dataset in all three cell types was approximately nine

amino acids with a standard deviation of approximately one, so a kappa CDR-L3 region of 11 amino acids or longer would be very unusual.

There was slightly more evidence of selection in the lambda repertoire than the kappa. The IGLV1 family significantly decreased in frequency as the B cells matured whereas the IGLV2 family increased. This appeared to be in part driven by an increase in frequency of the IGLV2-23 gene as B cells matured. Differences were also seen in the IGLJ genes with IGLJ1 increasing in frequency and IGLJ3 decreasing in frequency as the B cells matured (Figure 4-1B, page 194). There was also slightly more evidence of selection of the lambda CDR-L3. Lambda CDR-L3 with a high aliphatic index decreased in frequency as the B cells matured (Figure 4-3B, page 197), whereas lambda CDR-L3 with high isoelectric points (pI) and high instability index increased in frequency as the B cell matured, indicating that they may be promoted at tolerance (Figure 4-3D&F respectively).

It was interesting that there was more evidence of selection in the lambda repertoire compared to the kappa repertoire. It is known that the kappa locus tends to be rearranged before the lambda locus during receptor editing and lambda light chains have been associated with rescuing autoreactive heavy chains. One possibility may be that lambda light chains are more often paired with autoreactive heavy chains than kappa and these antibodies are more commonly selected out of the repertoire.

4.7.2 Kappa and Lambda Light Chains have Different CDR Properties

In agreement with previous findings (DeKosky et al., 2016), we found that lambda CDR-L3 are significantly longer and more hydrophobic than kappa CDR-L3 (Figure 4-4A, page 199). Given that long, hydrophobic CDR-H3 have been associated with autoreactivity (Wardemann et al., 2003) and are removed from the repertoire at tolerance checkpoints (Raaphorst et al., 1997, Volpe & Kepler, 2009, Larimore et al., 2012), it was surprising to find that lambda light chains have longer, more hydrophobic CDR-L3 than kappa as lambda light chains are rearranged after kappa during receptor editing (Brauninger et al., 2001) and are thought to be more able to rescue autoreactive heavy chains (Wardemann et al., 2004). This might suggest that the CDR-H3 and CDR-L3 have different roles in the antigen-binding site and that although long, hydrophobic CDR-H3 appear to be associated with autoreactivity, this may not be the case for long, hydrophobic CDR-L3.

Both kappa and lambda light chains had a high frequency of CDR-L3 with an isoelectric point (pI) of between 6.08 and 6.10. This suggests that a slightly acidic CDR-L3 pI (net negative charge at neutral pH) is advantageous for both kappa and lambda light chains. Conversely, very

few light chains of either class had CDR-L3 with a pI of 7.0, indicating that a net neutral charge at neutral pH is not desirable. It has been suggested that, in the heavy chain repertoire, CDR-H3 with basic pI (net positive charge at neutral pH) may be associated with polyreactivity (Zhang & Yeh, 1994). As lambda CDR-L3 have a lower mean pI and a higher frequency of acidic amino acids than kappa CDR-L3, it might be that the aforementioned ability of lambda light chains to rescue autoreactive heavy chains (Wardemann et al., 2004) is in some part due to an acidic CDR-L3 pI counteracting a basic CDR-H3 pI. Light chains with acidic CDR pI and high aspartic acid usage have been shown to be good at rescuing anti-DNA antibodies (Li et al., 2001, Kalinina et al., 2014). This may be due to the fact that DNA is negatively charged.

However, despite the speculations above that kappa and lambda light chains may complement certain heavy chains, the CDR-H3 physicochemical properties of heavy chains paired with kappa or lambda light chains showed no significant differences across all three donors, although there were some small donor-specific significant differences for some CDR-H3 properties (Figure 4-7A, page 204). We also found no significant difference in VDJ family usage between heavy chains paired with kappa or lambda light chains (Figure 4-7B). This analysis does not suggest that heavy chains with particular properties are predisposed to pair with either kappa or lambda light chains. This is in agreement with most of the literature on the topic which concludes that heavy-light chain pairing is random (DeKosky et al., 2016, Glanville et al., 2009, Brezinschek et al., 1998), however small biases in heavy-light chain pairing have been reported in some individuals (Jayaram et al., 2012). Despite these findings, it is possible that we are looking in the wrong place for heavy-light chain pairing bias. Biases may exist outside of the CDR loops, for example at the heavy-light chain interface.

Alternatively, the reason for the difference between kappa and lambda light chains may be independent of their heavy chain partners. Light chains are expressed by B cells in excess of heavy chains and therefore free light chains (also known as Bence Jones proteins) are expressed by plasma cells and can be found in the blood. There is some evidence that free light chains may have a role of their own in the immune response as it has been shown that they can bind antigen and activate mast cells in the absence of heavy chains (Thio et al., 2012). If free light chains do have a significant role in the immune system, it may be in this role that the difference between kappa and lambda is most significant.

Another hypothesis for the difference between kappa and lambda light chains is not that they have evolved to bind different types of antigen in the immune response, but that kappa and lambda light chains have evolved to respond to antigen in different ways. Due to the codons used in kappa and lambda variable regions, the two classes have quite different affinity

maturation patterns. Kappa CDRs are more prone to non-conservative mutations than lambda CDRs, however this means that they are more prone to the introduction of stop codons (Hershberg & Shlomchik, 2006). This means that the properties of kappa CDRs can change more dramatically more quickly than lambda CDRs, which change more gradually. The pattern of somatic hypermutation has been shown to be different between lambda light chains and heavy chains, suggesting that different chains use SHM in different ways (Boursier et al., 2003).

The physicochemical differences between kappa and lambda CDR-L3 were encoded in the germline (Figure 4-5, page 201) and are therefore the result of more than 450 million years of evolution since the divergence of the two classes (Greenberg et al., 1993). Mammals have two light chain classes (kappa and lambda), however, two light chain classes are not universal. Amphibians and bony fish have three classes (kappa, lambda and sigma), whereas birds have only the one (lambda) (Das et al., 2008, Das et al., 2012). Camelids have antibodies which lack light chains completely (Hamerscasterman et al., 1993). This indicates that a light chain is not essential for a functioning antibody. However, camelid antibodies lacking a light chain have alternative mechanisms of diversity, indicating that the additional diversity brought by light chains is advantageous (Nguyen et al., 2000).

In addition to physicochemical differences in the kappa and lambda CDRs, we also found differences in the structural composition of the CDR loops between the two classes. We found significant differences in the structural composition of kappa and lambda CDR-L3. In particular, we found that the likelihood of Beta structures was significantly higher in kappa CDR-L1 and CDR-L2, but significantly lower in CDR-L3 in comparison to lambda CDRs. The differences in CDR-L1 and CDR-L2 beta content was compensated for in lambda light chains by Helix and Coil propensities respectively (Figure 4-9A, page 206). There were significant differences between kappa and lambda CDR-L3 for Beta, Coil and Turn content. It has previously been reported that approximately 80% of kappa CDR-L3 have the same canonical structure, whereas lambda CDR-L3 have much more variety (Chailyan et al., 2011). As with the physicochemical properties, we found almost no structural differences in the CDR loops of heavy chains paired with kappa or lambda light chains (Figure 4-9B).

4.7.3 Random N-Nucleotide Addition is Positively Correlated within Individuals

One of the most striking findings of this chapter was the strong positive correlation in mean N-nucleotide addition at the kappa, lambda and heavy chain CDR3 regions within individuals (Figure 4-10, page 207). This suggests that there may be differences between individuals in the efficiency of the TdT enzyme or in the expression of the three TdT splice variants (hTdTS, hTdTL1 and hTdTL2). The three splice variants are expressed at various times during B cell and

T cell development and have differing exonuclease and transferase activity (Thai & Kearney, 2004).

Long CDR-H3 have been associated with autoreactivity (Wardemann et al., 2003) and it has been speculated that autoreactivity may be more likely when the CDR-H3 contains more non-germline encoded amino acids as they have not been subject to natural selection. Our analysis shows that some individuals can have on average three more N-nucleotides per CDR-H3 than others. This means that some people may have on average one more non-germline encoded amino acid per CDR-H3 than others. It would be interesting to investigate whether people with a higher mean N-nucleotide addition are more prone to autoimmune disease than those with less N-nucleotide addition or whether people with differing mean N-nucleotide addition respond to disease in different ways.

Chapter 5

The Antibody Response against Ebola

Chapter 5: The Antibody Response against Ebola

5.1 Introduction

Ebola is a devastating disease with a very high mortality rate; the outbreak in West Africa in 2014, caused by the Zaire strain of Ebola virus (EBOV), resulted in 28,616 cases and 11,310 deaths (WHO, 2016). On the 1st August 2018, the World Health Organisation (WHO) declared an outbreak of Ebola virus disease in the Democratic Republic of the Congo. At the time of writing, 78 cases had been identified, 44 of which were fatal (WHO, 2018b).

The virology of, and immune response to, Ebola are not fully understood. The early replication sites of Ebola are thought to be dendritic cells, monocytes and macrophages and the virus later disseminates to hepatocytes, fibroblasts and epithelial cells (Falasca et al., 2015). It is thought that Ebola invokes a very strong immune response which damages host tissues, resulting in a high mortality rate. However, some people do not suffer this outcome. The differences between people who survive and people who do not are not well understood. However it has been observed that a strong antibody response is a good indicator of survival (Ksiazek et al., 1999, Baize et al., 1999).

The primary target for anti-Ebola antibodies is the glycoprotein (GP) which coats the viral membrane and is used for viral entry into host cells. The membrane-bound GP is a trimer composed of three GP1-GP2 heterodimers and the structure is heavily glycosylated to shield the most conserved parts of the protein from the immune response (Figure 1-13, page 60). Some of the promising anti-Ebola drugs currently in development are monoclonal antibody cocktails targeting GP, for example ZMapp (see Figure 1-14, page 62, for the binding sites of well-known monoclonal antibody cocktails). A significant drawback of the antibody cocktails which have been used so far in Ebola outbreaks is that they are only specific to the Zaire strain of Ebola virus (EBOV). Although EBOV was responsible for the 2014 outbreak and is thought to be the strain with the highest mortality rate, the threat posed by other, lesser-studied, *Filoviridae* species to cause lethal outbreaks is not to be underestimated (Burk et al., 2016). However, cross-specific or cross-neutralising antibodies which bind GP from multiple *Filoviridae* species (particularly from more than one viral genus) are rare. It is important to keep searching for such antibodies as they may be valuable therapeutically either alone or as part of an antibody cocktail.

By examining the anti-GP B cell repertoire specifically, it has previously been reported that the anti-Ebola response may be very polyclonal and anti-GP antibodies show no clear bias in their

V(D)J usage or CDR3 properties (Bornholdt et al., 2016b). However, the whole B cell repertoire of recent Ebola survivors has not yet been studied. Therefore, in an effort to further our understanding of the features of a successful anti-Ebola antibody response, we used high-throughput sequencing of the B cell repertoires of recent Ebola survivors to investigate how they differ from negative, uninfected controls in terms of variable region properties, clonal expansions and heavy chain class usage.

Two high-throughput sequencing (HTS) datasets, known as the Ebola Response datasets (section 2.2, page 82) were built from survivors of the 2014 West Africa EBOV outbreak and also uninfected negative controls. The first HTS dataset, known as the Caucasian Aid Worker Survivor dataset, was built from total RNA isolated from the blood of three Caucasian aid workers and three uninfected negative controls. The second HTS dataset, known as the African Survivor dataset was built from total RNA isolated from the blood of five African EBOV survivors at two time points post-recovery. This dataset was useful for comparison with the Caucasian Aid Worker Survivor dataset to determine whether differences seen between the aid worker survivors and the negative controls were supported by trends seen in the African survivors.

The second aim of this chapter was to attempt to find new cross-reactive anti-GP antibodies using ribosome display. It has been shown previously that cross-reactive antibodies can be isolated from the repertoires of people who were infected with a single species of *Ebolavirus* (BDBV) (Flyak et al., 2016). Therefore, we hypothesised that we may be able to isolate new cross-reactive anti-GP antibodies using ribosome display libraries built from the B cell repertoires of recent EBOV survivors (see section 1.12.2.2, page 69, for the principle of ribosome display). The libraries were built using RNA from the same three Ebola survivors that were used for the Caucasian Aid Worker Survivor HTS dataset. The ribosome display libraries underwent one round of panning selection against EBOV GP and the specificity of the isolated scFv was then tested against GP from EBOV and other viruses.

In summary, the aims of this chapter were as follows:

1. Compare the B cell repertoire of recent EBOV survivors to negative controls to identify potential EBOV-specific features of the B cell repertoire
2. Find new anti-EBOV GP-specific antibodies using ribosome display and determine if they are cross-reactive to GP from multiple *Filoviridae* species
3. Investigate the selection pressures of the ribosome display and the properties of the GP-specific scFv that are isolated

5.2 B cell Repertoire Analysis of Caucasian Aid Worker Ebola Survivors

The template switch method of high-throughput sequencing was used to create the Caucasian Aid Worker Survivor dataset (see section 2.2, page 82, for method). This method of library construction added a UID to each cDNA transcript which allowed more reliable identification of clonal expansions than the method used to construct the B cell Development dataset (section 2.1, page 74).

The Caucasian Aid Worker Survivor dataset was built from mRNA isolated from whole blood samples taken from three Caucasian aid worker EBOV survivors and three healthy Caucasian age- and gender-matched negative controls who had never been infected with EBOV but did receive a yellow fever (YF) vaccine. Blood was taken from the negative controls at a baseline before receiving the vaccine (day 0) and again 28 days after receiving a yellow fever vaccine (day 28). This resulted in three sample groups:

1. Ebola survivors — Donor IDs: 215, 216, 217
2. YF vaccine controls (day 0) — Donor IDs: 193, 199, 207
3. YF vaccine controls (day 28) — Donor IDs: 193, 199, 207

The use of blood taken pre- and 28 days post- vaccination with the live, attenuated yellow fever vaccine allowed the distinction between B cell repertoire differences caused by viral infection in general and differences caused specifically by Ebola virus disease. 28 days is approximately the same as the time between Ebola infection and blood-taking in the Ebola survivors (although it was not possible to get the exact timings of this).

V(D)J gene usage and CDR3 physicochemical properties were calculated as described in section 2.4.2 (page 94). Clonotype clustering was performed as described in section 2.4.3 (page 95). Unless otherwise stated, only a single, modal, sequence from each clonal expansion was included in the data analysis so as not to skew the dataset with large clonal expansions.

V(D)J family/gene usage, CDR3 physicochemical properties and the size and class-composition of clonal expansions were investigated to determine whether there were any significant differences in the B cell repertoire of the Ebola survivors compared to the YF vaccine controls at day 0 and day 28.

5.2.1 V(D)J Gene Usage

The mean frequency (+/- 1 SD) of each heavy and light chain V(D)J family and gene was calculated for the IgA, IgG, IgM, IgK and IgL repertoires of each donor. The heavy chain dataset contained 24,415 entries in total and the light chain dataset contained 25,918 entries in total (see Table A-23 and Table A-25, pages 276-277, for the number of sequences per donor). Significant differences in V(D)J family usage between sample groups were identified using two-way ANOVA followed by Tukey's multiple comparisons post-test. Data were paired by donor (age-/gender-matched) and heavy chain (IgA, IgG, IgM) or light chain (IgK, IgL) class.

When investigating heavy chain IGHV gene and IGHD gene usage, some donor-class combinations had too few sequences for reliable frequency analysis and so were excluded from the heavy chain gene analysis (see section 2.5.1, page 96, for the "5X rule"). The samples excluded from IGHV gene and IGHJ gene analysis are shown in Table 5-1. Paired analysis for IGHV gene and IGHJ gene usage was not possible and so ANOVA followed by Tukey's multiple comparisons post-test was used on unpaired data. There were enough light chain samples to perform paired two-way ANOVA followed by Tukey's multiple comparisons post-test for light chain family and gene analysis (Table 5-1).

Table 5-1: Number of sequences in the Aid Worker Ebola Survivor V(D)J analysis. Samples excluded from IGHV gene analysis due to the "5X rule" are shown in red. Samples excluded from IGHD analysis are underlined in red. No sequences were excluded from the light chain analysis.

		Heavy Chain Classes			Light Chain Classes	
Group	Donor	IgA	IgG	IgM	IgK	IgL
Ebola Survivors	215	2910	3821	1745	3992	1420
	216	2009	2257	1427	3606	1195
	217	554	1177	1132	3663	406
YF Vaccine Controls Day 0	193	369	388	435	2133	347
	199	597	322	324	1964	339
	207	231	289	519	1800	291
YF Vaccine Controls Day 28	193	291	778	1106	1543	756
	199	310	273	118	1223	325
	207	308	126	599	443	472

Figure 5-1 shows the mean frequency (+/- 1 SD) of each heavy chain IGHV, IGHD and IGHJ family in the three sample groups. IGHV1 and IGHV3 families were significantly less common in the Ebola survivors than the YF vaccine day 0 and day 28 controls, whereas IGHV4 was significantly more common. IGHD3 family usage was significantly more common in the Ebola survivors than the YF vaccine control day 0 samples, but not the day 28 samples. There was no significant difference in the frequency of any of the IGHJ genes. Nine IGHV genes showed a significant difference in frequency between donor cohorts. IGHV1-2, IGHV1-69, IGHV1-8,

IGHV3-23, IGHV3-30, IGHV3-9 and IGHV4-38-2 were all significantly less frequent in the Ebola survivors than the YF vaccine controls. IGHV1-3 and IGHV4-39 were significantly more common in the Ebola survivors than the YF vaccine controls. IGHV4-39 in particular was a lot more frequently used in the Ebola survivors with a mean frequency in the B cell repertoire of 0.136 compared to 0.027 and 0.039 in the YF vaccine day 0 and day 28 controls respectively.

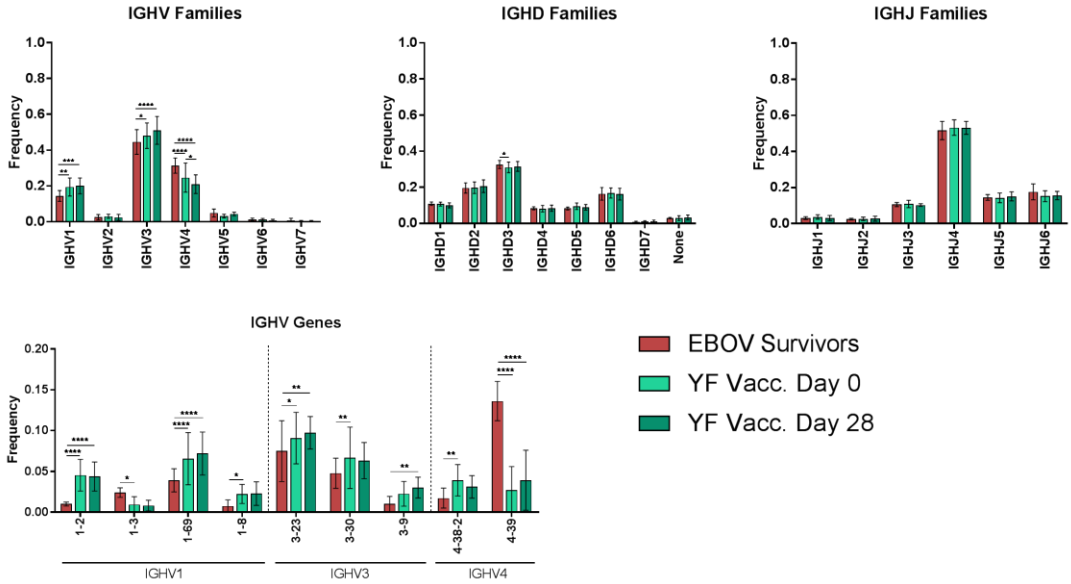


Figure 5-1: Heavy chain VDJ usage in the Caucasian Aid Worker Survivor dataset. The mean frequency of usage (+/- 1 SD) of all heavy chain VDJ families and also IGHV genes with a significant difference in frequency between donor groups are shown. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Figure 5-2 shows the mean frequency (+/- 1 SD) of VJ family and V gene usage in the kappa and lambda light chain repertoires. Figure 5-2A shows that none of the IGKV families differed significantly in frequency between the Ebola survivors and the YF vaccine controls. However, the IGKV genes IGKV1-39 and IGKV4-1 were significantly more frequent in the Ebola survivors than the YF vaccine controls. IGKJ1 was significantly more frequent in the Ebola survivors than the YF vaccine controls at day 28, but there was no significant difference in frequency with day 0. Figure 5-2B shows that there were no significant differences in IGLV family usage between the sample groups. However, the IGLV gene IGLV1-40 was significantly less frequent in the Ebola survivors than the YF vaccine controls and IGLV2-23 was significantly more frequent. IGLJ1 was significantly more frequent in the Ebola survivors than the YF vaccine controls and IGLJ3 was significantly less frequent.

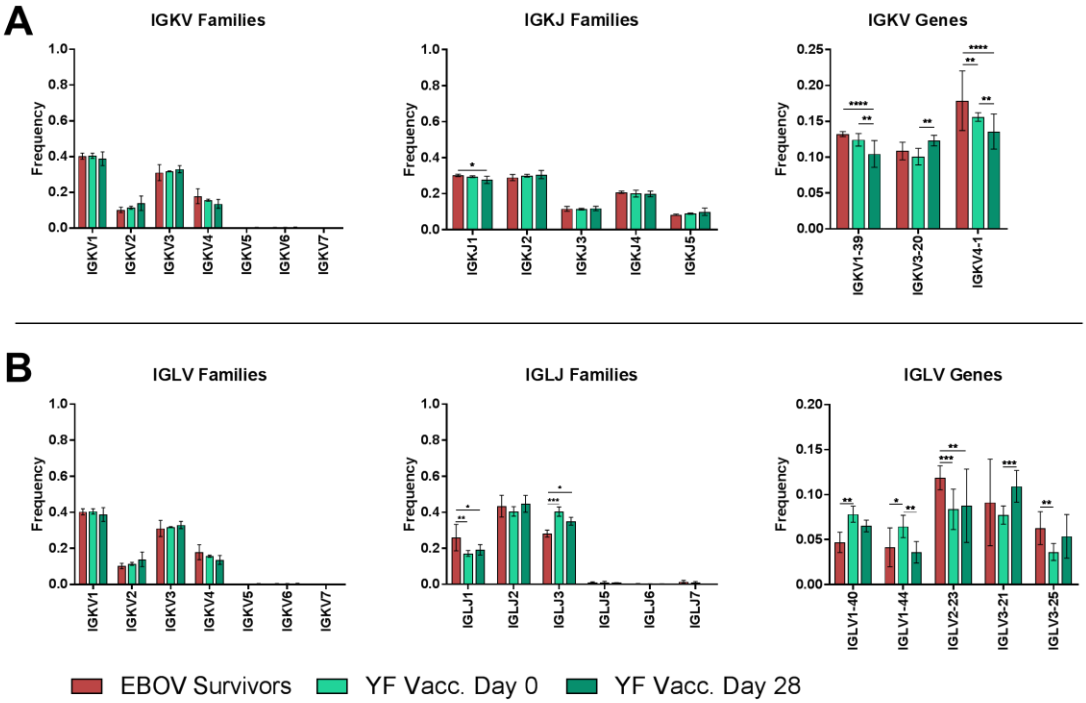


Figure 5-2: Light chain VJ usage in the Caucasian Aid Worker Survivor dataset. The mean frequency of usage (+/- 1 SD) of all (**A**) kappa and (**B**) lambda light chain V and J families and also V genes with a significant difference in frequency between donor groups are shown. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

5.2.2 CDR3 Physicochemical Properties

Before analysing the physicochemical properties of the heavy and light chain CDR3 regions, sequences which may skew the analysis were removed from the full dataset (section 2.5.2, page 96). The resulting heavy chain dataset contained 19,168 entries (Table A-24, page 276) and the resulting kappa and lambda datasets contained 15,582 and 4,505 entries respectively (Table A-26, page 277).

For each CDR3 amino acid sequence, the following physicochemical properties were calculated (section 2.4.2, page 94):

1. CDR3 length (amino acids)
2. Aliphatic index
3. Boman index
4. Isoelectric point
5. Hydrophobicity (GRAVY index)
6. Instability index

Data for each sample group (Ebola survivors, YF vaccine controls day 0 and YF vaccine controls day 28) were pooled and histograms were drawn for each heavy and light chain class (IgA, IgG,

IgM, IgK and IgL). Significant differences between the sample groups were identified using the Kruskal-Wallis test followed by Dunn's multiple comparisons post-test.

We found no significant differences in the heavy and light chain CDR3 physicochemical properties between the Ebola survivors and both YF vaccine control day 0 and day 28 time points. However, in both the kappa and lambda light chain repertoires, the mean CDR-L3 isoelectric point was significantly lower in the Ebola survivors than in the day 28 YF vaccine controls (7.11 vs. 7.21 in the kappa repertoire and 5.61 vs. 5.77 in the lambda repertoire).

5.2.3 Clonal Expansions

The heavy chain clone size distribution in each sample is shown in Figure 5-3 and the frequency of each class is also indicated. Only clonal expansions up to size 10 were plotted as clones larger than 10 were rare, however the largest clone size observed in each sample is indicated at the end of each x axis. Only heavy chain clonal expansions were investigated as public light chains make the identification of clonal expansions using light chains less reliable.

The most frequent clone size observed in all donors was a clone size of one. A clone size of one meant that no other related heavy chain was identified in the dataset. Approximately 80% of clones were of clone size one in all but one of the samples (YF vaccine control 199 day 0) which had a clone size one frequency of 62.6%. The remaining sequences were part of expanded clones (clone size two or bigger).

In the Ebola survivors, the majority of clone size one heavy chains were IgG (48.5%), IgA was the second most common (28.4%) and IgM was least common (23.2%). In Ebola survivor expanded clones the frequency of IgM classes occurred at a frequency of 3.5%, whereas IgA and IgG occurred at a frequency of 46.4% and 50.1% respectively.

Much greater variation was seen in the frequency of heavy chain classes of the YF vaccine controls. In donor 207 day 0 and day 28 and donor 193 day 28 the majority of clone size one sequences were IgM. However this was not seen in the other three YF vaccine control samples.

The variation seen in the frequency and class composition of clonal expansions in the YF vaccine controls may be due to the fact that fewer sequences were isolated from these samples.

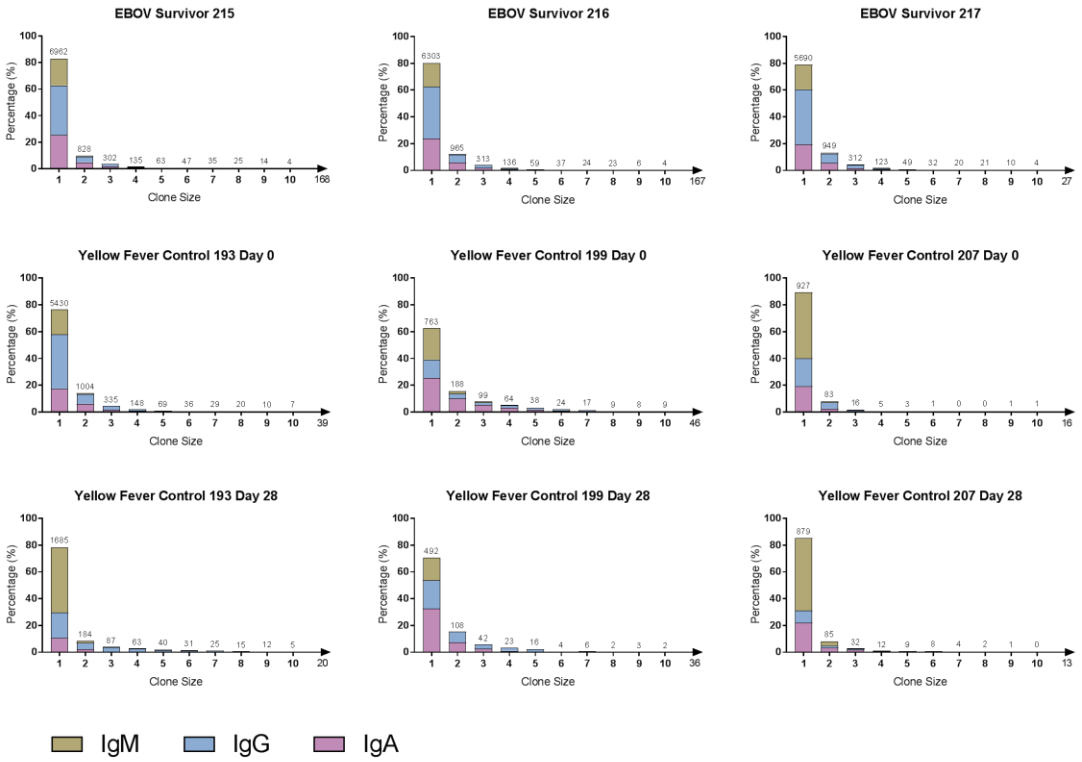


Figure 5-3: Clonal expansion frequency in Caucasian Aid Worker Survivor dataset. Only clone sizes up to 10 have been plotted as clones larger than this are rare. The largest clone size per donor has been indicated by an arrow on the x axis of each graph. The isotype frequency is also indicated. The number of clones of each size is indicated above the relevant bars.

5.2.4 Cross-Class Clones

It is often assumed that a clonal expansion consists of antibodies all of the same class (an IgA response, IgG response, IgM response etc.). Figure 5-3 was created using the isotype of the modal sequence of each clonal expansion and therefore each clonal expansion was displayed as a single class.

However, when examining all of the constituent sequences of each clonal expansion, we found that not every member of a clone did have the same heavy chain class. Some clonal expansions contained multiple heavy chain classes; we named such clonal expansions “cross-class clones”. We refer to clonal expansions containing a single heavy chain class as “single-class clones”. Table 5-2 shows the number of expanded clones identified in the Ebola Survivors and YF vaccine controls.

Table 5-2: Number of expanded clones in the Caucasian Aid Worker Survivor dataset. Expanded clones are those with a clone size of two or greater.

Sample	Donor	Single-Class Clones			Cross-Class Clones	Total Number of Clonal Expansions
		IgA	IgG	IgM		
Ebola Survivors	215	575	579	63	297	1514
	216	377	277	36	77	767
	217	63	179	42	51	335
YF Vaccine Controls Day 0	193	128	127	12	7	274
	199	280	147	38	15	480
	207	29	73	9	1	112
YF Vaccine Controls Day 28	193	55	359	58	18	490
	199	78	120	1	10	209
	207	81	32	40	1	154

The frequency of cross-class clones as a proportion of all expanded clones (clone size of two or greater) was calculated for each donor/time point. The frequency of cross-class clones was significantly higher (Fisher's exact test) in every Ebola survivor repertoire compared to every age- and gender-matched YF vaccine control at both time points (Figure 5-4A). The frequency of cross-class clones was between 10-20% in the Ebola survivors and between 1-3% and 1-5% in the day 0 and day 28 YF vaccine controls respectively.

Figure 5-4B shows the frequency of each heavy chain class combination in cross-class clones. The most frequent cross-class clone composition was IgA and IgG. The least frequent was IgA, IgG and IgM. The combination of IgA and IgM was slightly more common than IgM and IgG. Due to the small number of cross-class clones in the YF vaccine control samples, it was not possible to meaningfully compare the class composition of cross-class clones between the Ebola survivors and the YF vaccine controls.

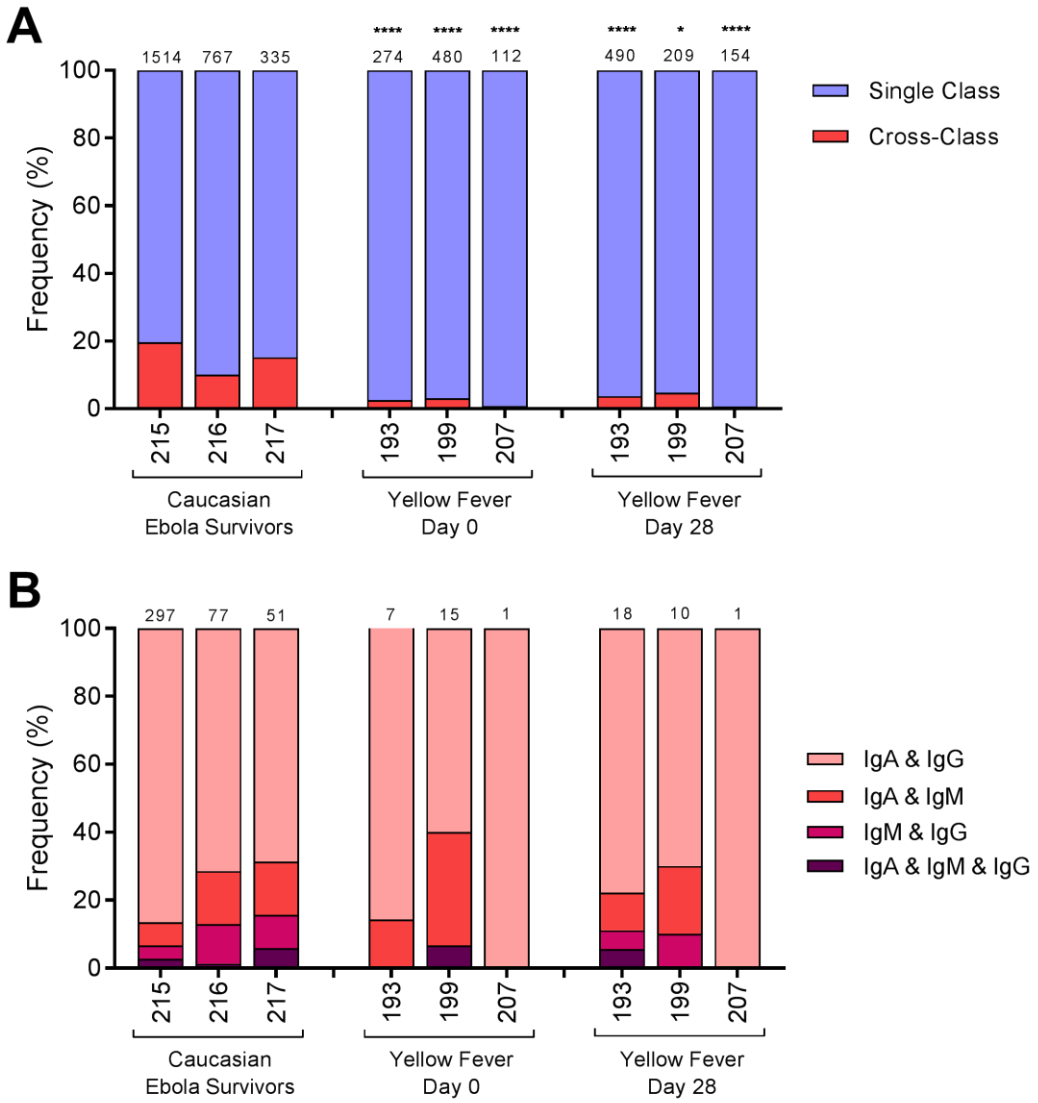


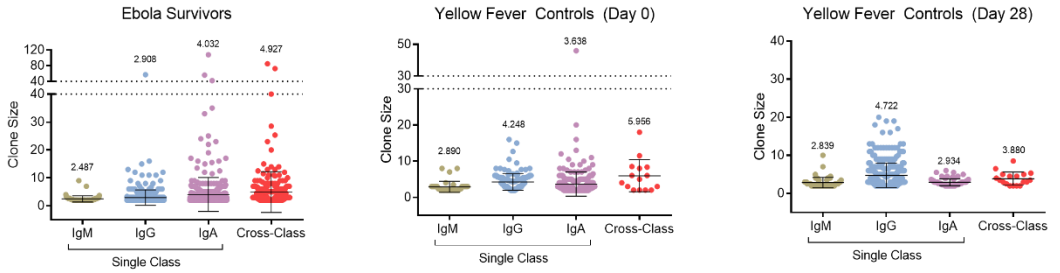
Figure 5-4: Frequency of cross-class clones in the Caucasian Aid Worker dataset. (**A**) The proportion of expanded clones which are single-class and cross-class. (**B**) The class combinations of each cross-class clone. The number of clones is displayed above each bar. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Multiple heavy chain classes within clonal expansions indicate that cross-class clones may have undergone multiple class-switch recombination events and therefore may have undergone significant clonal expansion and somatic hypermutation. We hypothesised that cross-class clones may therefore be on average larger than single-class clones and their variable regions may have more somatic hypermutation than those of single-class clones.

To investigate the size of cross-class clonal expansions compared to single-class clonal expansions, the mean sizes (+/- 1 SD) of all expanded clones were compared between the cross-class clones and single-class clones (separated by class) using the Kruskal Wallis test followed by Dunn's multiple comparisons post-test. Figure 5-5 shows that in the Ebola survivors, the mean cross-class clone size was significantly higher than IgM, IgG and IgA single-

class clones (4.927 compared to 2.487, 2.908 and 4.032 respectively). By contrast, the mean size of cross-class clones in the YF vaccine controls at day 0 was only significantly higher than the IgM single-class clones and in the YF vaccine control day 28 samples there was no significant difference in mean clone size between the cross-class clones and any of the classes of single-class clone.

The mean size of IgM single-class clones was approximately the same across all three sample groups. However, the mean size of the IgG and IgA single-class clones was very different between the Ebola survivors and the YF vaccine controls. In the Ebola survivors, the mean size of IgA single-class clones was significantly higher than IgG single-class clones. However, in both the YF vaccine control time points, the mean size of IgG single-class clones was significantly higher than IgG single-class clones.



Comparison	p value	Significance
IgM vs. IgG	0.1875	ns
IgM vs. IgA	<0.0001	****
IgM vs. Cross-Class	<0.0001	****
IgG vs. IgA	<0.0001	****
IgG vs. Cross-Class	<0.0001	****
IgA vs. Cross-Class	0.0045	**

Comparison	p value	Significance
IgM vs. IgG	<0.0001	****
IgM vs. IgA	0.1627	ns
IgM vs. Cross-Class	0.0051	**
IgG vs. IgA	<0.0001	****
IgG vs. Cross-Class	>0.9999	ns
IgA vs. Cross-Class	0.0987	ns

Comparison	p value	Significance
IgM vs. IgG	<0.0001	****
IgM vs. IgA	>0.9999	ns
IgM vs. Cross-Class	0.0623	ns
IgG vs. IgA	<0.0001	****
IgG vs. Cross-Class	>0.9999	ns
IgA vs. Cross-Class	0.3755	ns

Figure 5-5: Mean size of single-class and cross-class clonal expansions. The size of every clonal expansion (clone size of two or greater) is plotted and grouped by whether the clones are cross-class or single-class (IgM, IgG and IgA). The mean clone size (+/- 1 SD) in each group is plotted over the top and the mean clonal expansion size is written above each group.

The fact that, in the Ebola survivors, the mean clone size of cross-class clones was significantly higher than all three single-class clones, whereas this was not always the case in the YF vaccine controls presents the possibility that, as there were more sequences available from the Ebola survivors, more large clones could be identified and large clones have a higher probability of containing more than one heavy chain class than small clones. Therefore, to confirm that the Ebola survivors have a higher frequency of cross-class clones, independent of clone size, we measured the frequency of single-class clones and cross-class clones in the two most common expanded clone sizes: clone size two and clone size three (make up 52.8% and 19.4% of the expanded clones dataset respectively). In clone sizes two and three, the frequency of cross-class clones was still significantly higher (Fisher's exact test) in the Ebola survivors than the YF vaccine controls in most cases, indicating that it is not just an effect of clone size (Figure 5-6).

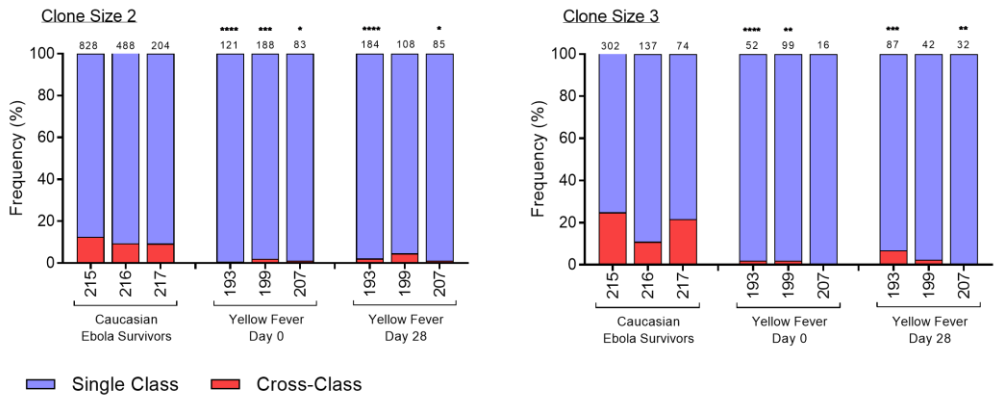
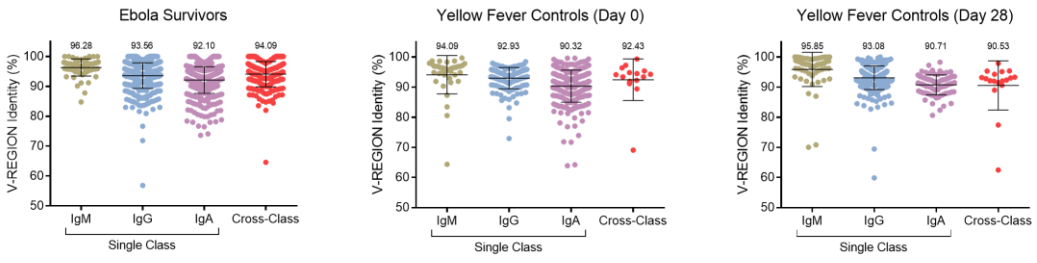


Figure 5-6: Frequency of cross-class clones in clone sizes two and three in the Caucasian Aid Worker Survivor dataset. The proportion of expanded clones that are cross-class and single-class are indicated and the number of clones in each group is displayed above each bar. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

The level of somatic hypermutation (SHM) of the modal sequence of each clonal expansion was measured using the IMGT HighV-QUEST V-REGION percentage identity; the lower the V-REGION percentage identity, the more SHM had taken place. Significant differences in V-REGION percentage identity were measured using the Kruskal Wallis test followed by Dunn's multiple comparisons post-test. Figure 5-7 shows that in the Ebola survivors, cross-class clones had the same level of V-REGION identity as IgG single-class clones (94.09% and 93.56% respectively) and significantly higher V-REGION identity than IgA single-class clones (92.10%). As expected, IgM single-class clones had significantly higher V-REGION identity than all other groups (96.28%). At both YF vaccine control time points, the V-REGION identity was highest in the IgM single-class clones and lowest in the IgA single-class clones. The V-REGION identity of the cross-class clones was not significantly different to the IgG or IgA single-class clones in either YF vaccine control time point (Figure 5-7).



Comparison	p value	Significance
IgM vs. IgG	<0.0001	****
IgM vs. IgA	<0.0001	****
IgM vs. Cross-Class	<0.0001	****
IgG vs. IgA	<0.0001	****
IgG vs. Cross-Class	0.7106	ns
IgA vs. Cross-Class	<0.0001	****

Comparison	p value	Significance
IgM vs. IgG	0.0461	*
IgM vs. IgA	<0.0001	****
IgM vs. Cross-Class	>0.9999	ns
IgG vs. IgA	<0.0001	****
IgG vs. Cross-Class	>0.9999	ns
IgA vs. Cross-Class	0.0747	ns

Comparison	p value	Significance
IgM vs. IgG	<0.0001	****
IgM vs. IgA	<0.0001	****
IgM vs. Cross-Class	0.0001	***
IgG vs. IgA	<0.0001	****
IgG vs. Cross-Class	>0.9999	ns
IgA vs. Cross-Class	0.7009	ns

Figure 5-7: V-REGION identity of clonal expansions. The V-REGION percentage identity to germline of the single, modal, sequence of each single-class (IgM, IgG and IgA) and cross-class clonal expansion. The mean V-REGION percentage identity (+/- 1 SD) is plotted over the top. The mean V-REGION percentage identity is written above each group.

To determine whether the properties of cross-class clone variable regions differed from single-class clone variable regions, the heavy chain VDJ family frequencies and CDR-H3 properties were compared. Due to the small number of expanded clones, there were too few sequences to reliably investigate the frequency of IGHV and IGHD genes (see section 2.5.1, page 96, for information on the "5X rule"). Only the Ebola survivors were investigated as there were too few cross-class clones in the YF vaccine controls to reliably investigate them.

Figure 5-8 shows the mean frequency (+/- 1 SD) of each heavy chain VDJ family in the single-class and cross-class clones. The data were paired by donor and two-way ANOVA followed by Tukey's multiple comparisons post-test was used to identify significant differences between the types of clonal expansion. There were no significant differences in IGHV, IGHD or IGHJ family usage between the cross-class clones and any of the single-class clones.

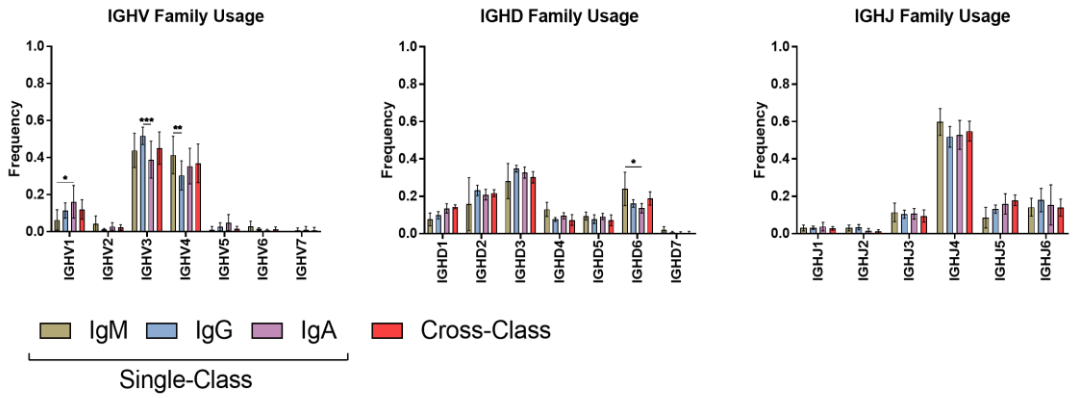


Figure 5-8: Heavy chain VDJ family usage in cross-class and single-class clones. The mean frequency (+/- 1 SD) of each IGHV, IGHD and IGHJ family in the cross-class and single-class clones of the three Ebola survivors. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

The CDR-H3 physicochemical properties (as listed on page 219) of the cross-class and single-class clones in the Ebola survivors were compared using the Kruskal-Wallis test followed by Dunn's multiple comparisons post-test. No significant differences in CDR-H3 physicochemical properties were found.

5.3 African Ebola Survivors also have high Prevalence of Cross-Class Clones

It was shown in section 5.2.4 that the Caucasian aid worker Ebola survivors had a significantly higher frequency of cross-class clones in their clonally-expanded B cell repertoire than the YF vaccine controls (Figure 5-4A). To determine whether this high frequency of cross-class clones is donor-specific or if it may be related to Ebola virus infection, we measured the frequency of cross-class clones in a second B cell repertoire dataset; the African Survivor dataset. The African Survivor dataset was built from total RNA isolated from the whole blood of five African survivors of the 2014 EBOV West Africa outbreak at two time points after their recovery (see section 2.2, page 82, for full materials and methods). Serology was available for the five donors (donor 14, 15, 2, 50 and 82) and showed varying IgM and IgG response to EBOV GP in the months following infection (Figure 2-4, page 84).

We found that the frequency of cross-class clones in the African EBOV survivors was very similar to that of the Caucasian aid worker EBOV survivors (Figure 5-9A). Furthermore, the frequency of cross-class clones was not significantly different (Fisher's exact test) at the two time points for four out of the five African survivors despite there being a gap of up to eight months between them.

The class composition of the cross-class clones varied a little between time points, but there was no clear pattern (Figure 5-9B) and there was no clear relationship with the serology patterns shown in Figure 2-4 (page 84). Like in the Caucasian aid worker data, the overwhelming majority of cross-class clones in the African survivors were composed of a combination of IgA and IgG antibodies.

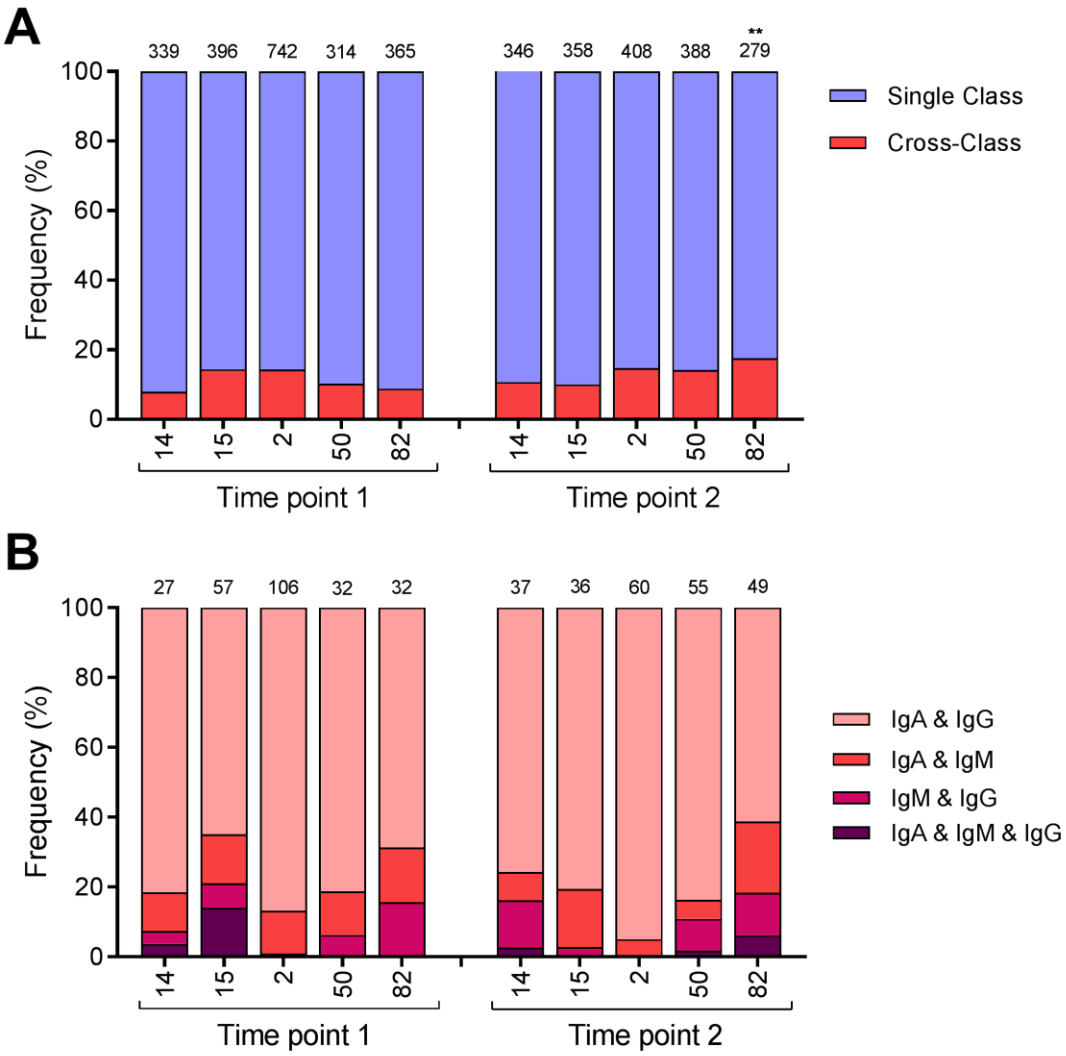
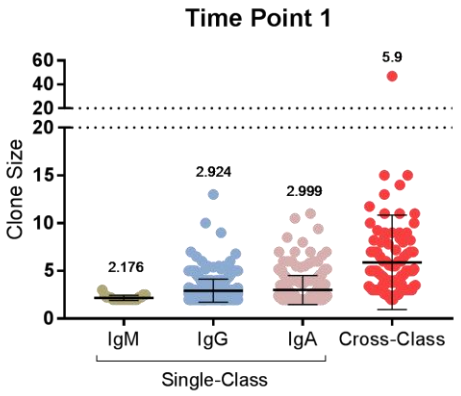


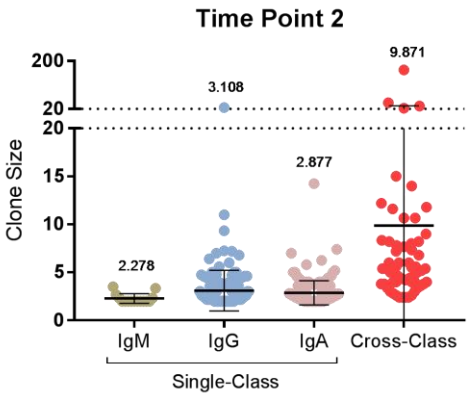
Figure 5-9: Frequency of cross-class clones in the African Ebola survivors. (**A**) The proportion of expanded clones which are single-class and cross-class. (**B**) The class combinations of each cross-class clone. The number of clones is displayed above each bar. The numbers on the x-axis are donor IDs. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

The cross-class clones in the Caucasian aid worker EBOV survivors were shown to have a slightly higher mean clone size than the single-class clones (Figure 5-5, page 224). In the African survivors, at both time points, the mean size of the cross-class clones was significantly higher than the single-class clones (Figure 5-10).

Like in the Caucasian aid worker survivors (Figure 5-7, page 225), the cross-class clones of the African survivors did not have very different V-REGION identities to the single-class clones (Figure 5-11).

In both cases, significance was measured using the Kruskal Wallis test followed by Dunn's multiple comparisons post-test.

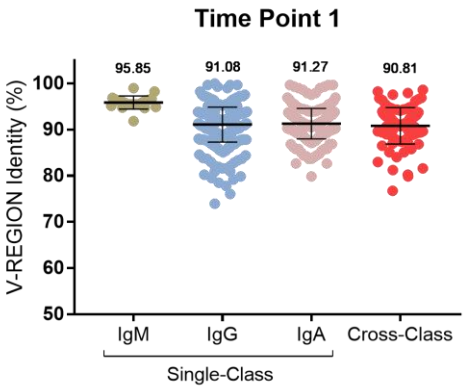


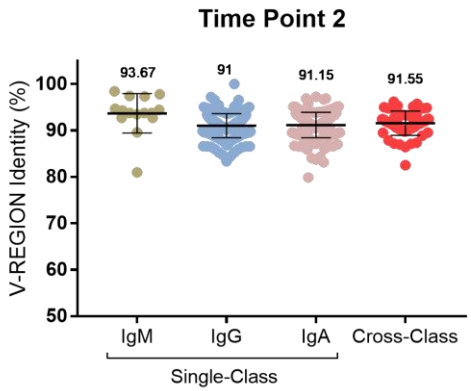


Time Point 1		
Comparison	p value	Significance
IgM vs. IgG	0.0139	*
IgM vs. IgA	0.0406	*
IgM vs. Cross-Class	<0.0001	****
IgG vs. IgA	>0.9999	ns
IgG vs. Cross-Class	<0.0001	****
IgA vs. Cross-Class	<0.0001	****

Time Point 2		
Comparison	p value	Significance
IgM vs. IgG	0.0442	*
IgM vs. IgA	0.0557	ns
IgM vs. Cross-Class	<0.0001	****
IgG vs. IgA	>0.9999	ns
IgG vs. Cross-Class	<0.0001	****
IgA vs. Cross-Class	<0.0001	****

Figure 5-10: Cross-class and single-class clone sizes in African Ebola survivors. Each point represents a clonal expansion. The mean clone size (+/- 1 SD) is plotted over the top.





Time Point 1		
Comparison	p value	Significance
IgM vs. IgG	<0.0001	****
IgM vs. IgA	<0.0001	****
IgM vs. Cross-Class	<0.0001	****
IgG vs. IgA	>0.9999	ns
IgG vs. Cross-Class	>0.9999	ns
IgA vs. Cross-Class	>0.9999	ns

Time Point 2		
Comparison	p value	Significance
IgM vs. IgG	0.0004	***
IgM vs. IgA	0.002	**
IgM vs. Cross-Class	0.0285	*
IgG vs. IgA	>0.9999	ns
IgG vs. Cross-Class	0.4719	ns
IgA vs. Cross-Class	>0.9999	ns

Figure 5-11: Cross- and single-class clone V-REGION identity in African Ebola survivors. Each point represents the single, modal, sequence in a clonal expansion. The mean V-REGION identity (+/- 1 SD) is plotted over the top.

5.4 Anti-EBOV GP scFv were Isolated using Ribosome Display

Ribosome display libraries were constructed using total RNA from the same three EBOV survivors used for the Caucasian Aid Worker Survivor HTS dataset (section 5.2). The principle of ribosome display is described in section 1.12.2.2 (page 69) and the complete materials and methods are set out in sections 2.14 and 2.15 (page 118 and 125 respectively). Briefly, a kappa and lambda PCR3 input library was constructed for each donor (six libraries in total). Each library was panned against EBOV GP (expressed in-house using the method described in section 2.17, page 132). The mRNA of binders was reverse-transcribed and the cDNA was PCR amplified (PCR4), creating PCR4 output libraries. The PCR4 output libraries were inserted into pESL expression vectors upstream of a NanoLuc luciferase gene (Figure A-2, page 271). The plasmids were transformed into *E. coli* and grown on selective agar. scFv binding to GP was tested by growing individual *E. coli* colonies in selective LB broth and using the scFv-containing clarified supernatant in a solid-phase ELISA. scFv binding was detected using a luciferase assay (see section 2.16, page 131, for full materials and methods).

Quality control experiments were required to confirm that the ribosome display method was a reliable way of isolating EBOV GP-specific scFv from the donor repertoires. The quality control experiments were:

1. Confirming that the diversity of the PCR3 input library was high
2. Confirming that scFv created using this protocol can bind antigen specifically

The results of the quality control experiments are described below in sections 5.4.1 and 5.4.2. The results of the panning selection are then described in sections 5.4.3 and 5.4.4.

5.4.1 Diversity of the Ribosome Display PCR3 Input Libraries

Unlike the high-throughput sequencing libraries (sections 5.2 and 5.3), the ribosome display libraries were built using many rounds of PCR and multiple V gene-specific primers. This introduced the risk of considerable over-/under-representation of certain sequences due to primer bias. Large biases may have led to libraries with a low diversity, which would risk potential binders being missing from the library.

It was therefore important to check the diversity of the PCR3 input libraries to confirm that they were representative of the donor repertoires. To do this, the PCR3 input libraries of donor 215 and 217 (there was not enough sample to sequence the donor 216 libraries) were sequenced using high-throughput sequencing (see section 2.3, page 92, for method). The clonal diversity, V(D)J frequency and CDR3 properties of the PCR3 input libraries were compared to the donor 215 and 217 HTS libraries (section 5.2). The HTS libraries were built

using the template-switch method with a universal primer site and had PCR duplicates removed from the library. The HTS libraries should therefore be highly representative of the donors' repertoires.

To compare the clonal diversity of the PCR3 input libraries and the HTS libraries, a diversity curve was drawn for the heavy and light chain sequences of each library (section 2.5.4, page 97). The diversity curves (Figure 5-12) show that the diversity of the PCR3 input libraries was similar to that of the HTS libraries, although the diversity of the PCR3 input libraries was slightly lower. However, this indicates that the PCR3 input libraries had good diversity and were representative of the donors' repertoires.

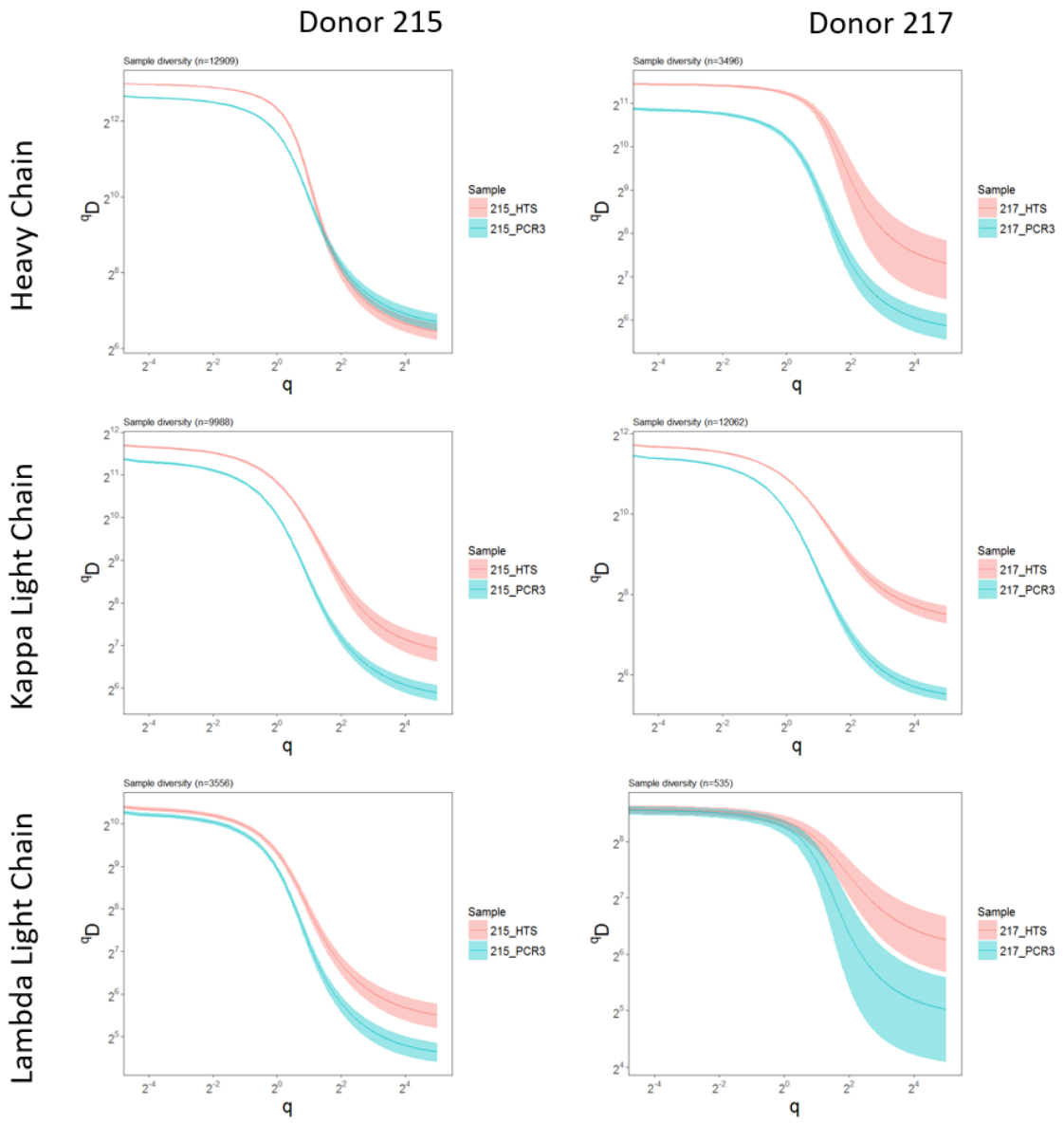


Figure 5-12: Diversity of HTS library and ribosome display PCR3 input library of donors 215 and 217. General diversity index (^{q}D) with 95% CI versus the clonal frequency (q). As q tends towards zero, rarer clones are weighted more heavily, as q tends towards infinity common clones are weighted more heavily. Lower ^{q}D values indicate lower diversity.

Next, the range of V(D)J families and CDR3 length distribution in the PCR3 input libraries was compared to the HTS datasets to confirm that they were representative of the donors' repertoires.

The mean heavy and light chain V(D)J family usage (+/- 1 SD) in donor 215 and 217 PCR3 input libraries and HTS libraries was plotted. Due to the small number of donors, statistical analyses were not meaningful in comparing V(D)J frequencies. Histograms were also plotted of the heavy and light chain CDR3 lengths in the donor 215 and 217 PCR3 input libraries and HTS libraries. CDR3 lengths of the PCR3 input library and HTS library were compared for each donor using the Kruskal-Wallis test followed by Dunn's multiple comparisons post-test.

The mean frequency of each heavy chain VDJ family was quite similar between the PCR3 input and HTS libraries, except IGHV3 was less frequent in the PCR3 input library than the HTS library. There was no significant difference in the distribution of CDR-H3 lengths between the PCR3 input libraries and the HTS libraries (Figure 5-13A).

The IGKV and IGKJ family usage was similar between the PCR3 input and HTS libraries. However, there were some differences in the IGLV and IGLJ libraries; IGLV2 and IGLJ2 were less common in the PCR3 input library and IGLV3 and IGLJ1 were more common. There was no significant difference in the kappa or lambda CDR-L3 length distribution between the PCR3 input libraries and the HTS libraries (Figure 5-13B).

Overall, this suggests that, despite the use of V gene-specific forward primers and multiple rounds of PCR, the diversity and representativeness of the PCR3 input library is quite similar to that of the HTS library. This suggests that the PCR3 input library has the diversity required for effective use in a panning selection.

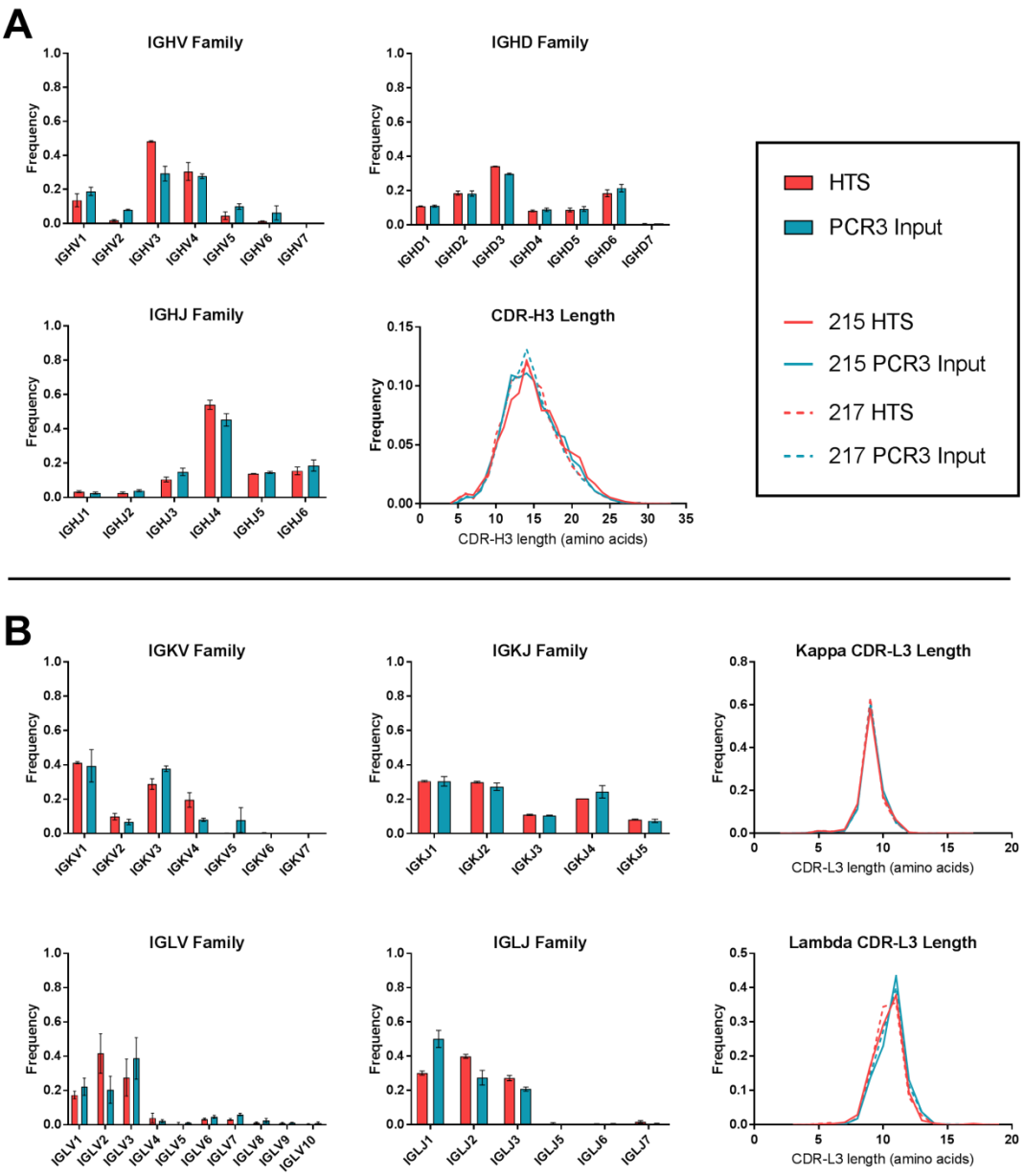


Figure 5-13: Diversity of PCR3 input library compared to HTS dataset. The mean frequency (+/- 1 SD) of each V(D)J family and a histogram of the CDR3 length distribution for the (**A**) heavy chain repertoire and (**B**) light chain repertoires in the ribosome display PCR3 input library and the HTS dataset.

5.4.2 Antigen-Specificity of scFv

scFv are composed of a heavy and a light chain variable region linked by a linker peptide (instead of the constant region) and are expressed by *E. coli* instead of mammalian cells. The scFv were expressed with a NanoLuc luciferase enzyme at the C-terminus (Figure 2-9, page 131). As the scFv are structurally different to antibodies and are expressed from *E. coli*, there was a risk that the scFv may not show specificity in the same way that the same variable regions would in antibody form. Therefore, to confirm that antibody binding specificity is maintained in scFv form, we expressed KZ52 (Maruyama et al., 1999) as an scFv and tested its specificity.

KZ52 is a well-known human antibody which binds EBOV GP at a conformational epitope which spans the GP1 and GP2 subunits (Lee et al., 2008) (Figure 1-14, page 62). The specificity of the KZ52 scFv was tested in a solid-phase ELISA against three viral glycoproteins: [100]

1. EBOV GP
2. Marburg virus Ravn strain (RAVV) GP
3. HIV gp120 dimers

As the KZ52 antibody is highly specific to EBOV GP, KZ52 scFv should only show binding to EBOV GP. As predicted, KZ52 scFv showed specific binding to EBOV GP and did not show binding to either RAVV GP or HIV gp120 (Figure 5-14). To confirm that only EBOV GP-specific scFv will bind EBOV GP, a negative control scFv (discovered during an early practice selection) was tested in solid-phase ELISA against the same glycoproteins. The negative control scFv showed no binding against any of the glycoproteins, confirming that only specific scFv will bind EBOV GP (Figure 5-14).

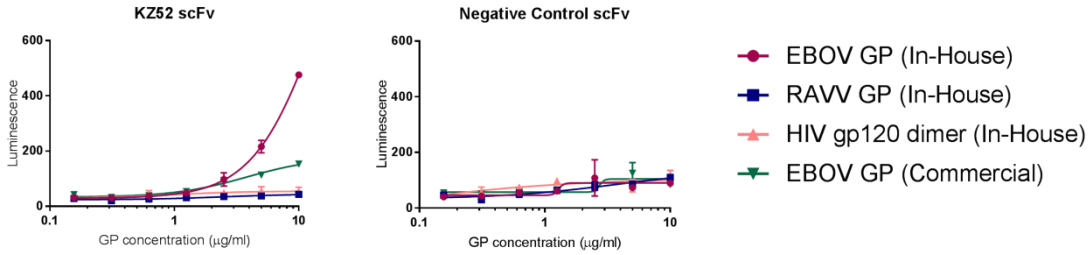


Figure 5-14: Positive and negative controls confirm scFv specificity. KZ52 binds specifically to in-house and commercial EBOV. The negative control scFv does not bind any of the glycoproteins. 50 µl neat *E. coli* supernatant was applied to every well. Supernatant was tested in duplicate and binding was detected by measuring luminescence (mean +/- 1 SD shown).

[100] EBOV GP, RAVV GP and HIV gp120 were all made in-house, however the scFv were also tested against a commercial EBOV GP (Zaire Ebolavirus envelope glycoprotein, P-300-100, Icosagen) to confirm that the scFv would also bind that

5.4.3 EBOV GP-Specific scFv Isolated from Panning Selection

After a single round of panning selection against in-house EBOV GP, 94 *E. coli* colonies from the kappa and lambda output libraries were picked per donor (188 colonies per donor in total) and screened for scFv inserts using colony PCR. Colonies containing scFv inserts were grown in 1 ml liquid culture and the scFv-containing clarified supernatants were used in solid-phase ELISA to test for binding to EBOV GP. Binding specificity was confirmed by testing 50 µl neat scFv-containing supernatants in triplicate against GP titrated in two-fold serial dilutions from 10 µg/ml to 0.156 µg/ml and against wells coated with blocking buffer only (3% (w/v) skimmed milk powder in PBS).

Of the 564 scFv isolated from the kappa and lambda libraries of the three Caucasian aid worker EBOV survivors, 22 EBOV GP-binding scFv were isolated. Seven were isolated from donor 215, six were isolated from donor 216 and nine were isolated from donor 217. However one was a duplicate of 217_A01 and was therefore excluded from the reported results.

Due to differing levels of scFv expression, the peak luminescence values varied between samples. However, because the luminescence from the NanoLuc-conjugated scFv decreased as the concentration of EBOV GP on the ELISA well decreased, this indicates that the scFv were binding specifically to EBOV GP (Figure 5-15).

Every scFv in Figure 5-15 was sequenced using Sanger sequencing. The heavy and light chain nucleotide sequences were analysed using IMGT V-QUEST to determine V(D)J gene usage, CDR3 amino acid sequence and V-REGION identity. The heavy chain class (IgA, IgG, IgM) of the original antibody was deduced from the linker sequence (Table 2-36, page 118). The sequencing information is shown in Table 5-3 (page 237). Four of the 21 sequences shown in Table 5-3 are non-functional either because they are out of frame or incomplete.

Seven of scFv heavy chain sequences in Table 5-3 appeared in the repertoire of their respective donors in the heavy chain Caucasian Aid Worker Survivor HTS dataset (215_E06, 215_F09, 215_G07, 216_B01, 216_D01, 217_A07, 217_A09). All but one of these sequences (215_G07) were part of expanded clones in the HTS dataset. Moreover, all of the sequences were either part of a cross-class clone in the HTS dataset (215_E06) or the scFv isolated with ribosome display was of a different IgH class to the clone identified in the HTS dataset, indicating that the clone is cross-class.

Only the seven scFv which appeared in the HTS dataset were taken forward for further investigation of their binding specificity in sections 5.5 and 5.6.

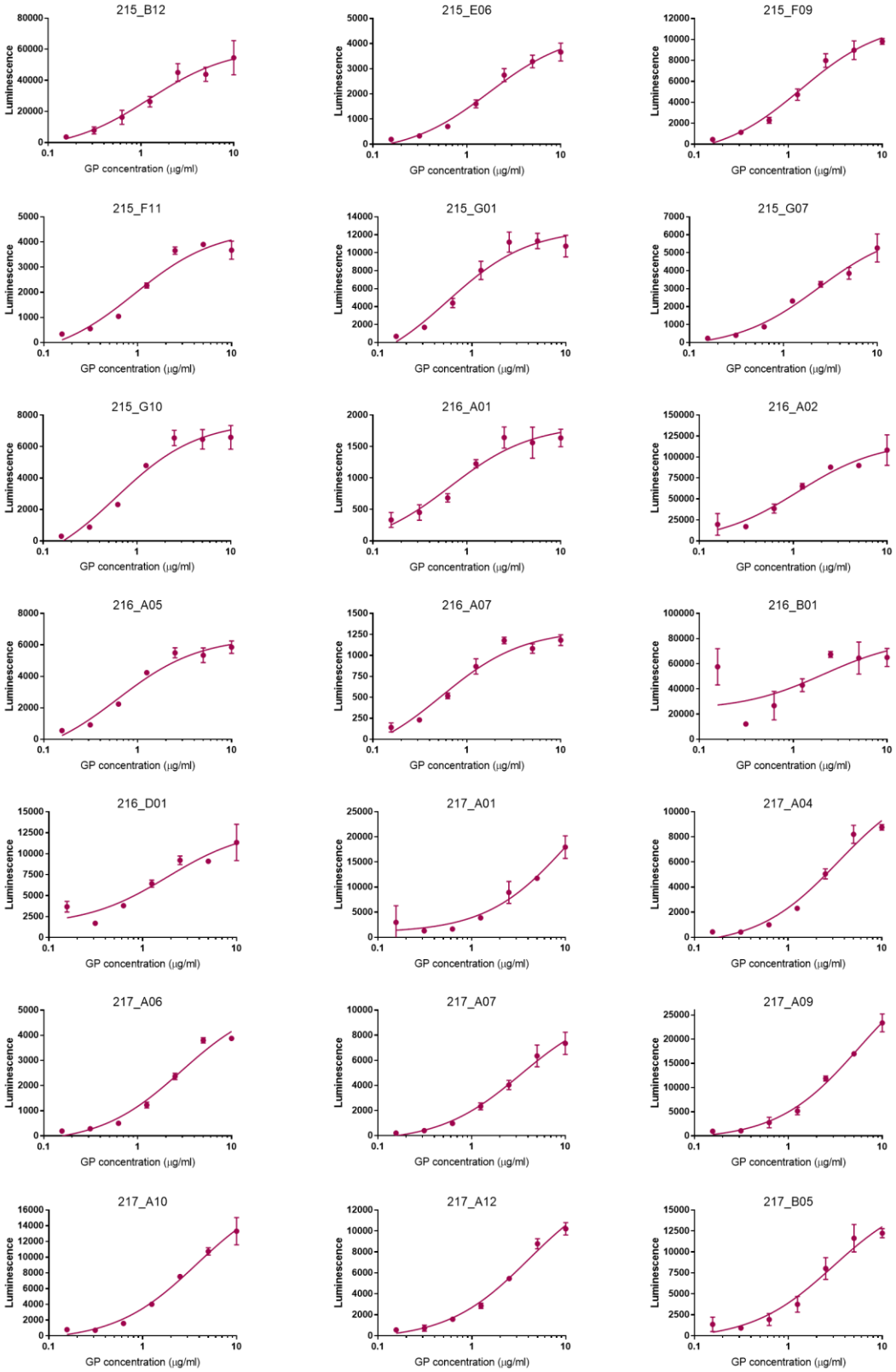


Figure 5-15: EBOV GP binding curves of scFv isolated using ribosome display. Mean luminescence (+/- 1 SD) of scFv tested in triplicate against EBOV GP in solid-phase ELISA. Neat scFv-containing *E. coli* supernatant was applied to each ELISA well. EBOV GP was titrated using six 1 in 2 dilutions in PBS from a starting concentration of 10 µg/ml.

Table 5-3: IgH and IgL details of EBOV GP-binding scFv. V(D)J and CDR3 as defined by IMGT V-QUEST. IgH class is determined by the scFv linker sequence. % V ID refers to the percentage V-REGION identity (i.e. identity to germline). HTS Dataset indicates whether an identical CDR-H3 amino acid sequence (IMGT definition) occurs in the HTS dataset.

	Heavy Chain						Light Chain					HTS Dataset		
scFv ID	V gene	D gene	J gene	CDR-H3	% V ID	Class	V gene	J gene	CDR-L3	% V ID	Class	CDR-H3 match?	Clone size	Clone class
215_B12	IGHV2-26	IGHD6-19	IGHJ5	ARIRQSGWEHWFDP	93.13	IgM	IGKV5-2	IGKJ1	LQHDNFPRT	96.77	Kappa	✗	-	-
215_E06	IGHV6-1	IGHD1-1	IGHJ4	ARSQDYFDY	94.61	IgA	IGKV5-2	IGKJ2	LQHDNFPYT	97.49	Kappa	✓	2	IgM, IgG
215_F09	IGHV2-26	IGHD6-19	IGHJ4	GRIRGRSQWLSHLDY	95.53	IgM	IGKV5-2	IGKJ2	LQHDNFPYT	98.57	Kappa	✓	2	IgA1
215_F11	IGHV3-11	IGHD3-3	IGHJ6	AGFWSGLQNTYGMDV	92.01	IgM	IGLV3-19	IGLJ2/3	HSRDSSGHPRVV	97.85	Lambda	✗	-	-
215_G01	IGHV6-1	IGHD3-3	IGHJ4	ARYDFWSNYGIFDY	96.97	IgG	IGLV3-19	IGLJ2/3	QSRDSGGNHLV	96.42	Lambda	✗	-	-
215_G07	IGHV4-34	IGHD1-26	IGHJ3	AKEGPDMTYYRHAFDI	91.93	IgM	IGLV1-44	IGLJ1	AAWDDNLNSFV	92.28	Lambda	✓	1	IgA2
215_G10	IGHV1-3	IGHD6-13	IGHJ3	ARRNRRGDRPQQLVGAAFDI	99.65	IgM	IGLV3-19	IGLJ2/3	NSRDSSGNHVV	98.92	Lambda	✗	-	-
216_A01	-	-	-	-	-	-	-	-	-	-	-	-	-	-
216_A02	IGHV6-1	IGHD3-9	IGHJ4	ARSFSWAFDY	96.63	IgM	IGKV1-16	IGKJ2	QQYNSYPYT	97.85	Kappa	✗	-	-
216_A05	IGHV3-74	IGHD6-19	IGHJ4	ARGPFGSGWTGFDY	97.57	IgA	IGKV5-2	IGKJ1	PQHENFRRS	97.85	Kappa	✗	-	-
216_A07	-	-	-	-	-	-	-	-	-	-	-	-	-	-
216_B01	IGHV4-59	IGHD3-22	IGHJ6	ARTFYESPGDFYNYYFFHMDV	90.53	IgM	IGKV3-20	IGKJ5	QHYGSSL	93.26	Kappa	✓	5	IgA1
216_D01	IGHV1-69	IGHD3-22	IGHJ4	ARAMLYYETSGYFPLAS	92.01	IgM	IGLV3-19	IGLJ2/3	KSRDSSGDRLV	93.55	Lambda	✓	10	IgA1
217_A01	IGHV3-73	IGHD3-22	IGHJ4	TRRSTDESRGYYTS	94.56	IgA	IGKV5-2	IGKJ2	LQHDNFPYT	99.28	Kappa	✗	-	-
217_A04	IGHV4-34	IGHD6-13	IGHJ4	ARGSGHGSSSSY	94.39	IgA	IGKV1-39/1D-39	IGKJ1	QQYNTYPVT	93.19	Kappa	✗	-	-
217_A06	-	-	-	-	-	-	-	-	-	-	-	-	-	-
217_A07	IGHV3-7	IGHD4-17	IGHJ4	ARVPYGDSLRYYDS	91.32	IgA	IGKV1-39	IGKJ4	QQSYSTPPLT	97.13	Kappa	✓	2 3	IgG1 IgG1/2
217_A09	IGHV6-1	IGHD5-12	IGHJ4	ARGTPVVLRTLVSGSGFDF	97.64	IgM	IGKV5-2	IGKJ2	LQHDNFPLYT	98.92	Kappa	✓	2	IgA1
217_A10	IGHV1-18	IGHD3-16	IGHJ6	AREATYGSEKYGTDV	90.62	IgM	IGKV1-6	IGKJ1	QHYGSSLWT	93.91	Kappa	✗	-	-
217_A12	-	-	-	-	-	-	-	-	-	-	-	-	-	-
217_B05	IGHV3-74	IGHD2-8	IGHJ5	VRVTGGWNWFDP	95.49	IgA	IGLV3-25	IGLJ2	QSADSSGTFVV	88.17	Kappa	✗	-	-

5.4.4 How Panning Selection Altered the Ribosome Display Libraries

To determine whether selection took place, the diversity of the PCR4 output libraries was compared to the diversity of the PCR3 input libraries. Figure 5-16 shows that the PCR4 output libraries are much less diverse than the PCR3 input libraries. This suggests that the PCR4 libraries were built from fewer starting sequences than the PCR3 library and therefore suggests that the panning selection reduced the diversity of the library by removing some of the sequences which did not bind EBOV GP.

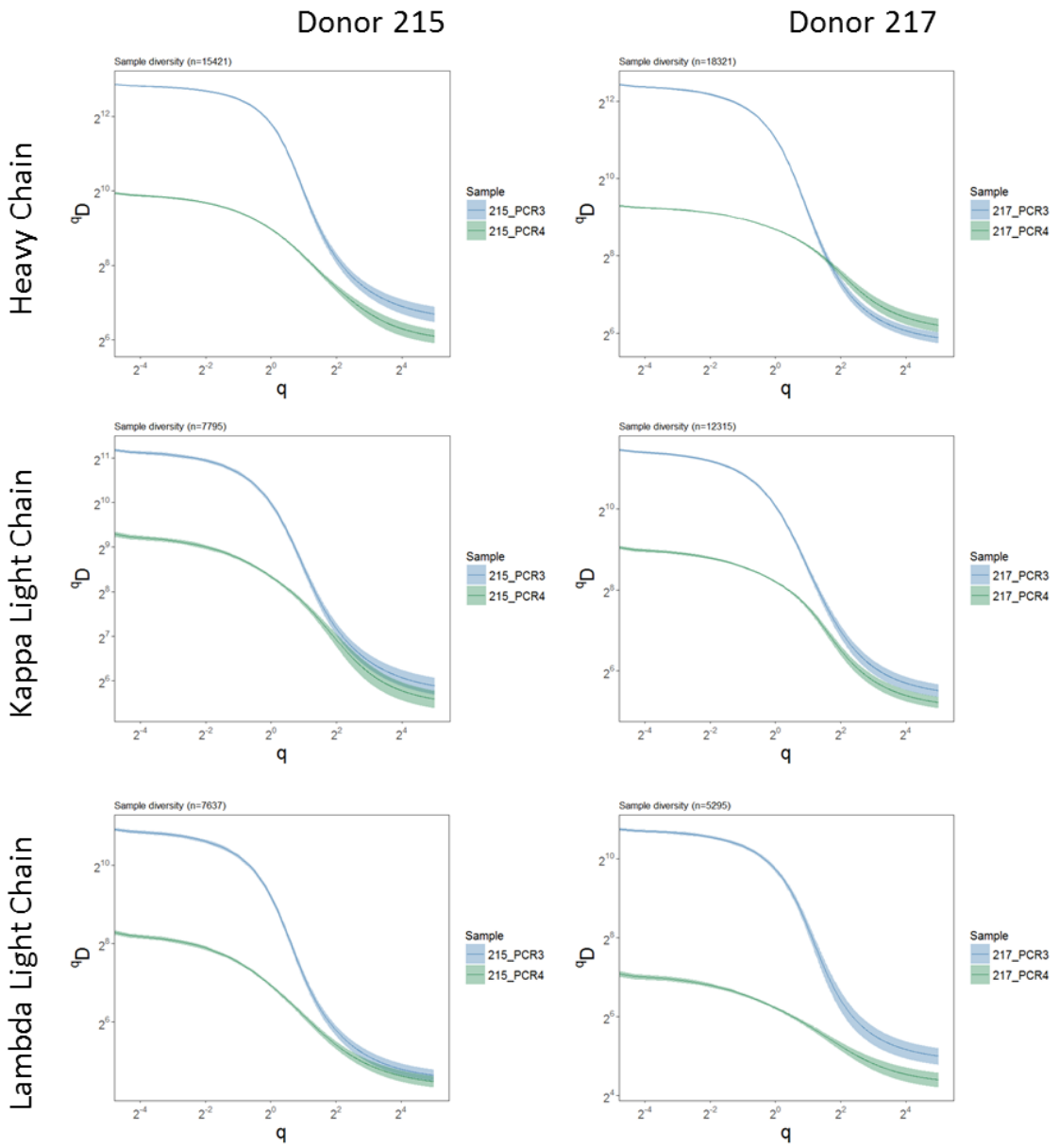


Figure 5-16: Diversity curves of PCR3 input and PCR4 output libraries. General diversity index (qD) with 95% CI versus the clonal frequency (q). As q tends towards zero, rarer clones are weighted more heavily, as q tends towards infinity common clones are weighted more heavily. Lower qD values indicate lower diversity. In every case, the PCR4 output libraries are much less diverse than the PCR3 input libraries.

Although Figure 5-16 shows that the heavy and light chain diversity decreased in the PCR4 output library, there was little evidence of directional selection in the heavy and light chain V(D)J family usage (Figure 5-17); only the frequency of some IGLV and IGLJ families was slightly different after panning selection. Furthermore, there were no significant differences in any of the CDR-H3 or CDR-L3 region properties between the PCR3 input and PCR4 output libraries.

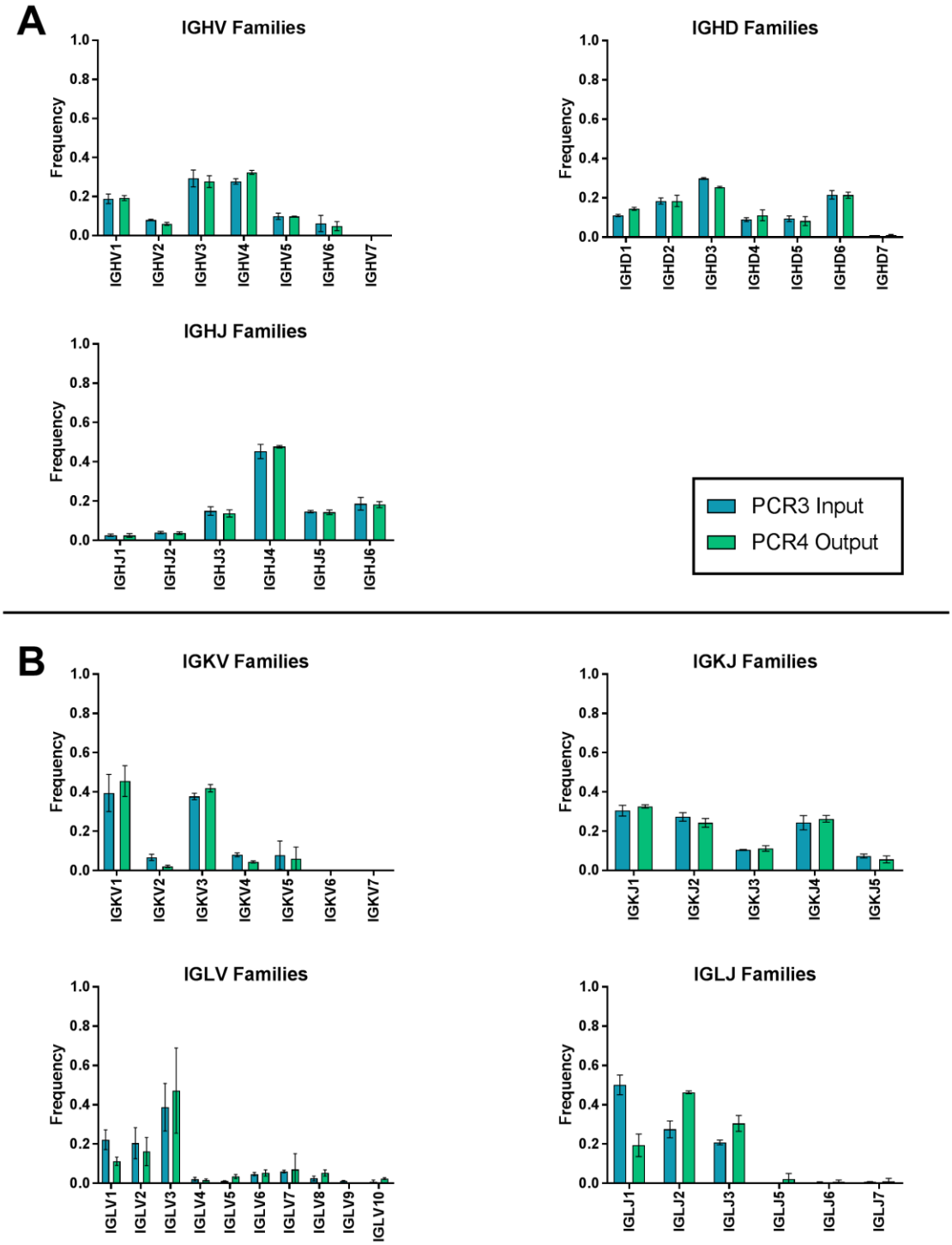


Figure 5-17: V(D)J family frequency in PCR3 input and PCR4 output libraries. The mean frequency (+/- 1 SD) of each (**A**) heavy chain and (**B**) light chain V(D)J family in donor 215 and 217 PCR3 input and PCR4 output libraries.

When looking at the V gene and D gene usage in the selected scFv (Table 5-3, page 237) some occur frequently (e.g. IGHV2-26, IGHV4-34, IGHV3-74, IGHD6-19, IGHD3-3, IGHD3-22, IGKV5-2 and IGLV1-19). Therefore, the frequency of these genes in the PCR3 input and PCR4 output libraries was plotted to investigate if they increased in frequency after panning selection, implying that they had been selected for. Figure 5-18 shows the mean frequency (+/- 1 SD) of each gene used in the selected scFv. There was little evidence of strong positive selection of any of the V genes or D genes, and the frequency of several genes decreased after panning selection (e.g. IGHV3-11, IGHD3-22, IGHD3-9, IGHD5-12 and IGKV3-20). However, IGHV1-69, IGHV4-34 and IGLV3-25 increased in frequency after the panning selection (Figure 5-18).

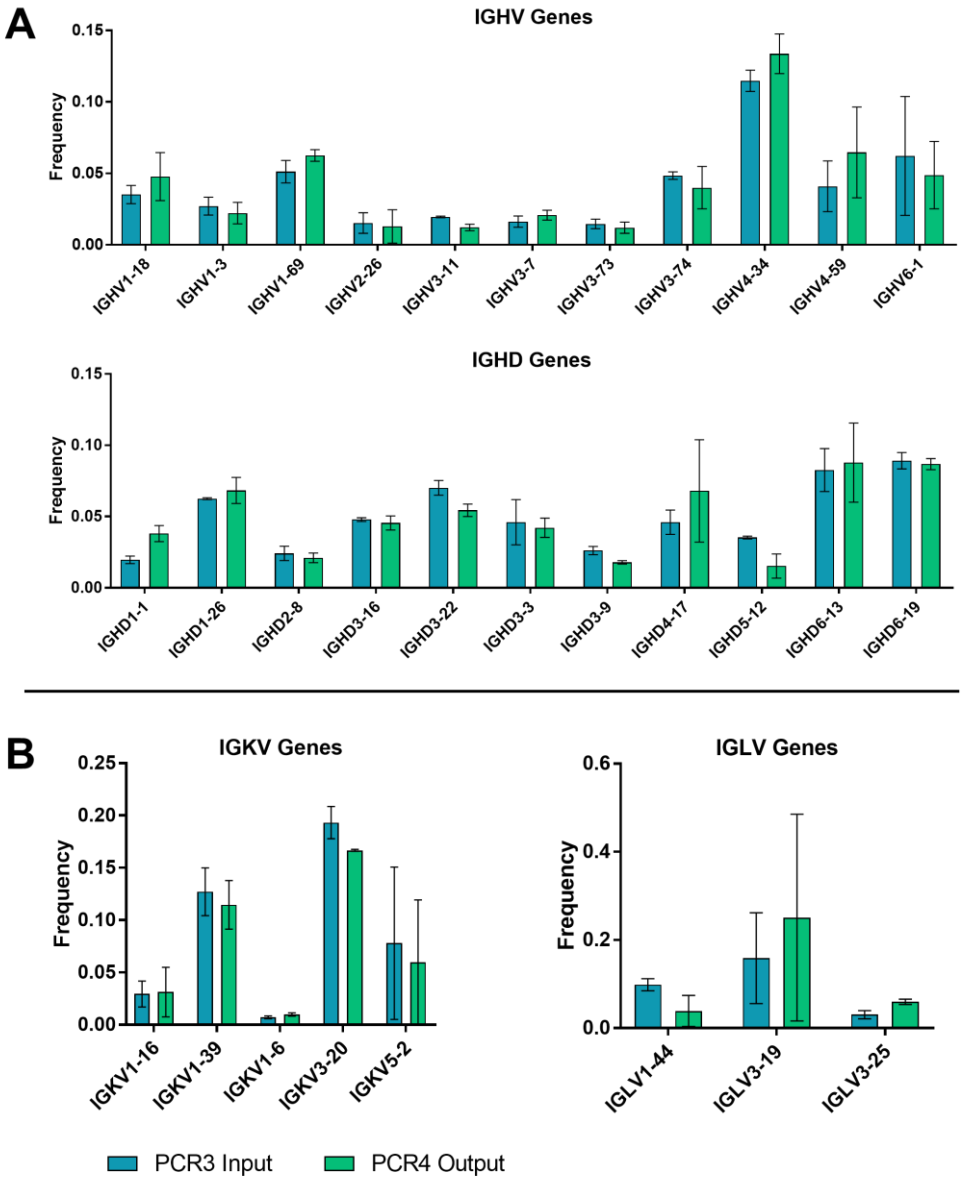


Figure 5-18: Frequency of scFv genes in PCR3 input and PCR4 output libraries. The mean frequency (+/- 1 SD) in the donor 215 and 217 PCR3 input and PCR4 output libraries of each (**A**) heavy chain and (**B**) light chain gene used by the scFv isolated by panning selection and listed in Table 5-3 (page 237).

5.5 scFv Cross-React with EBOV GP, RAVV GP and HIV gp120

Antibodies which bind GP from more than one *Filoviridae* species could be very useful therapeutically, but such antibodies are rare. The seven EBOV GP-binding scFv which occurred in the HTS dataset (Table 5-3, page 237) were tested for binding GP isolated from Marburg (Ravn) virus (RAVV GP) to determine whether they could cross-react with GP from other *Filoviridae*. As a negative control, the scFv were also tested for binding to HIV gp120 dimers.

It was hypothesised that if the scFv were specific to EBOV GP only, they would not bind RAVV GP or HIV gp120. If the scFv were cross-reactive with GP from multiple *Filoviridae*, they would bind EBOV GP and RAVV GP but not HIV gp120. Binding was assessed using solid-phase ELISA.

Surprisingly, the scFv bound all three glycoproteins (Figure 5-19). Even more surprisingly, they showed strongest binding to HIV gp120. This suggested that the scFv may be binding to something all the glycoproteins had in common. All the glycoproteins were glycosylated, and all of them were His-tagged.

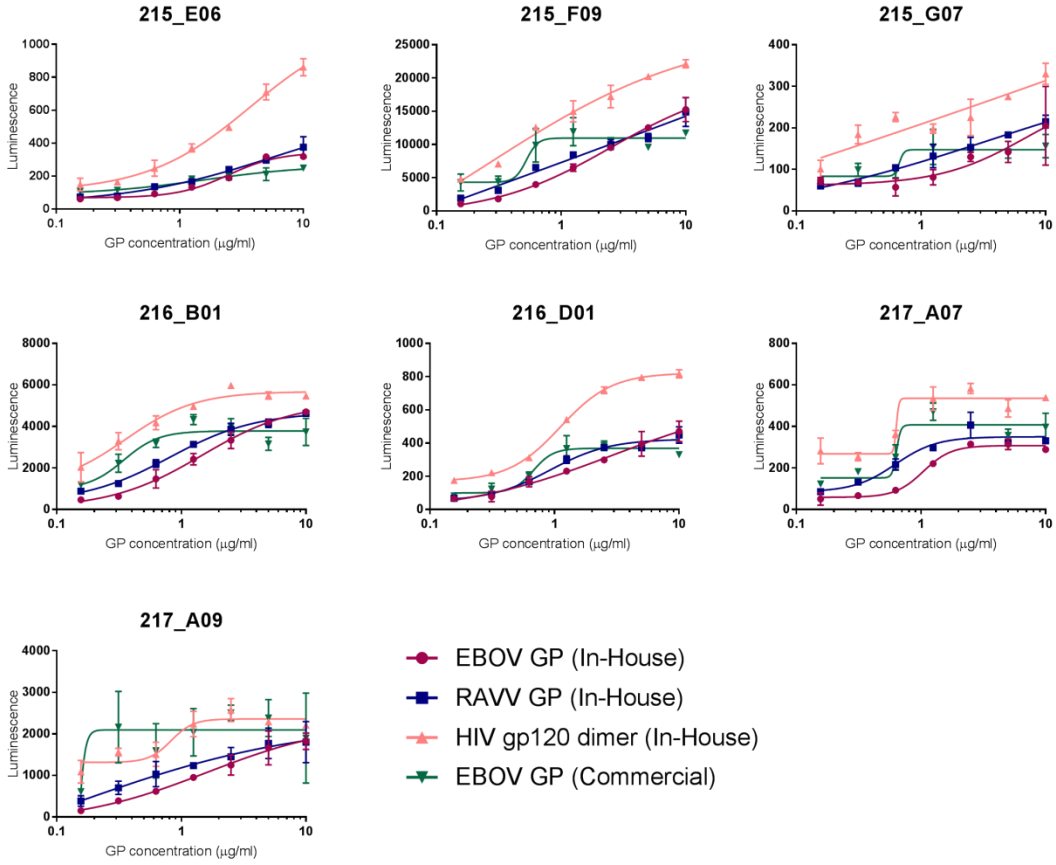


Figure 5-19: Specificity of scFv against three viral glycoproteins. Mean luminescence (+/- 1 SD) of scFv tested in duplicate against EBOV GP (both in-house and commercial), RAVV GP and HIV gp120 in solid-phase ELISA. Each glycoprotein was titrated using six 1 in 2 dilutions in PBS from a starting concentration of 10 µg/ml. 50 µl neat scFv-containing *E. coli* supernatant was applied to each ELISA well.

5.6 Glycoprotein-Binding scFv Bind Less to the Deglycosylated Form

All three glycoproteins investigated are heavily glycosylated. EBOV GP and RAVV GP have both N-linked and O-linked glycans, whereas gp120 has almost exclusively N-linked glycans. All three glycoproteins have a His-tag.

To investigate whether the seven scFv which showed binding to EBOV GP, RAVV GP and HIV gp120 (Figure 5-19) may be binding N-linked glycans, the glycoproteins were deglycosylated and tested again for scFv binding using solid-phase ELISA.

The glycoproteins were deglycosylated by incubating with the enzymes Endo H and PNGase F overnight at 37°C (see section 2.17.4, page 134 for method). Endo H is a recombinant glycosidase which cleaves the chitobiose core of high-mannose glycans but it does not cleave complex glycans. PNGase F is an amidase which cleaves the innermost N-acetylglucosamine and asparagine residues in both high-mannose and complex glycans. Both enzymes only cleave N-linked glycans (see Figure 5-20).

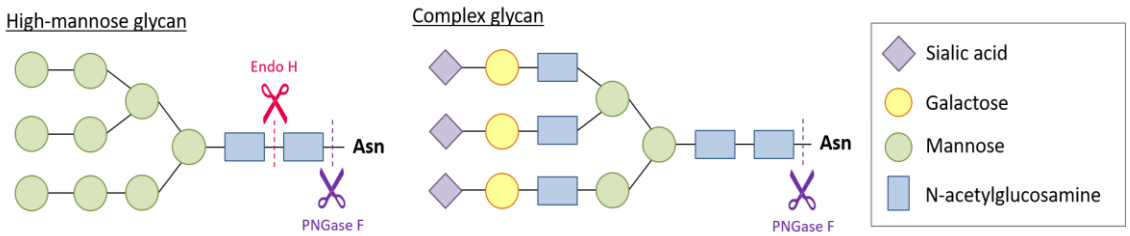


Figure 5-20: Endo H and PNGase F glycan cleavage sites. Endo H only cleaves the chitobiose core of high-mannose glycans. PNGase F cleaves just above the asparagine (Asn) residue of both high-mannose and complex glycans, therefore removing the whole glycan. It can sometimes improve the efficiency of PNGase F to "clear" some of the high-mannose glycans using Endo H, providing better access for PNGase F, hence why both enzymes were used.

The successful cleavage of EBOV GP, RAVV GP and HIV gp120 by Endo H and PNGase F was confirmed by SDS-PAGE (Figure 5-21A) and ELISA using biotinylated lectin PHA-L (Figure 5-21B). PHA-L binds mannose, galactose and N-acetylglucosamine on complex N-linked glycans and therefore the loss of PHA-L binding indicates deglycosylation. The integrity of the protein was confirmed by checking that KZ52 scFv binds both complete and deglycosylated EBOV GP (Figure 5-21C).

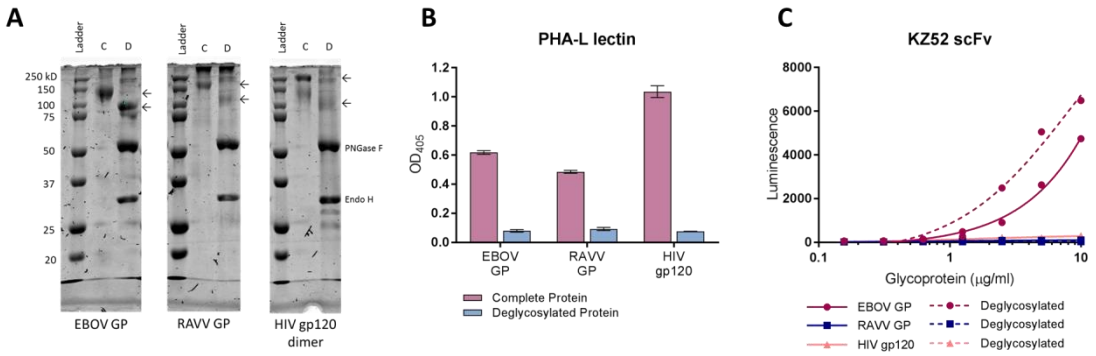


Figure 5-21: Confirmation of glycoprotein deglycosylation. (**A**) 10% SDS-PAGE stained with Coomassie Blue. Arrows indicate the position of the complete (C) and deglycosylated (D) glycoproteins. (**B**) Mean OD_{405} (+/- 1 SD) of biotinylated PHA-L lectin tested in triplicate against complete and deglycosylated glycoproteins in a solid-phase ELISA. Binding was detected using alkaline phosphatase-conjugated streptavidin. (**C**) Mean luminescence (+/- 1 SD) of scFv tested in duplicate against complete and deglycosylated EBOV GP, RAVV GP and HIV gp120 in solid-phase ELISA. Neat scFv-containing *E. coli* supernatant was applied to each ELISA well. Each glycoprotein was titrated using six 1 in 2 dilutions in PBS from a starting concentration of 10 µg/ml.

After confirming that EBOV GP, RAVV GP and HIV gp120 had been successfully deglycosylated, the seven scFv shown in Figure 5-19 (page 241) were tested for binding to the complete and deglycosylated forms of these three glycoproteins (Figure 5-22). 215_F09 and 216_B01 showed much greater binding to the complete glycoproteins than the deglycosylated glycoproteins, suggesting that these two scFv might be specific to the N-linked glycans. The expression of the other five scFv was not as good, however 215_E06, 215_G07 and 217_A09 showed somewhat more binding to the complete GP than to the deglycosylated GP, suggesting that they could be binding the N-linked glycans. 216_D01 and 217_A07 did not show a large difference in binding between the complete GP and deglycosylated GP, suggesting that they probably are not binding the N-linked glycans. However, this result is inconclusive as these two scFv had very low expression.

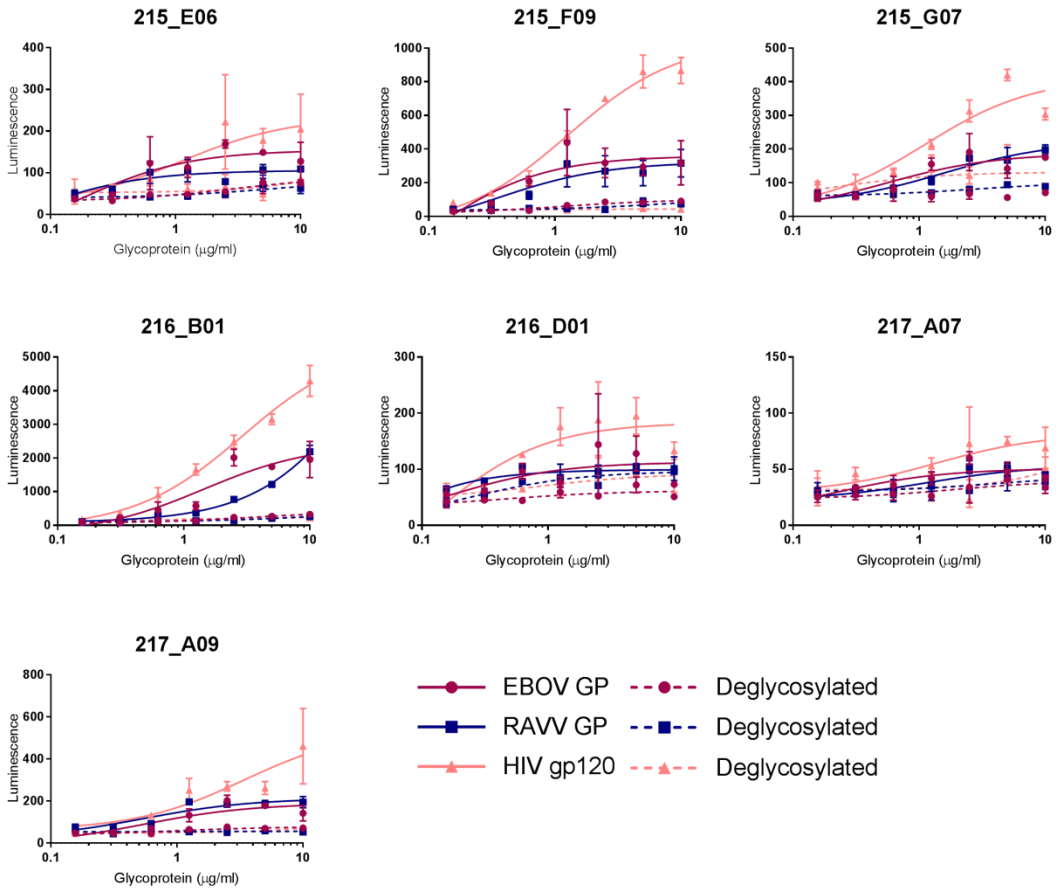
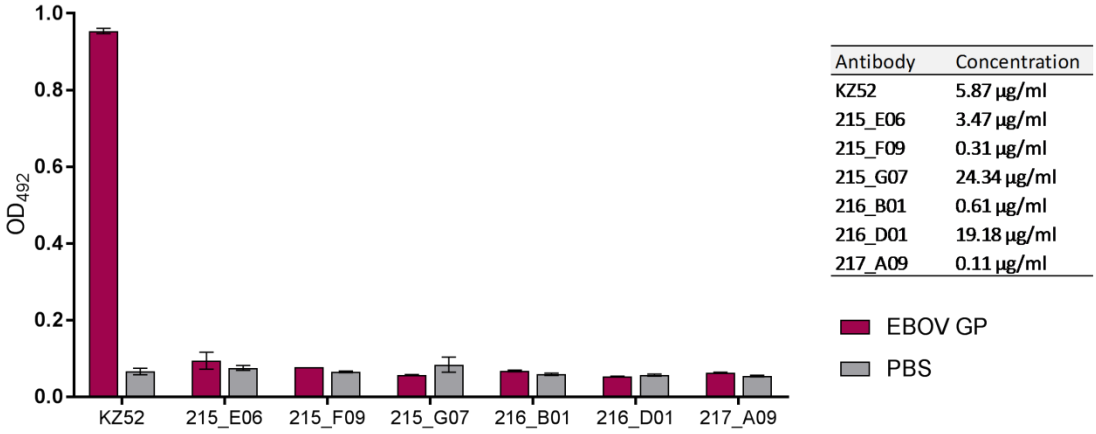


Figure 5-22: Specificity of scFv to complete and deglycosylated glycoproteins. Mean luminescence (+/- 1 SD) of scFv tested in duplicate against complete and deglycosylated EBOV GP, RAVV GP and HIV gp120 in solid-phase ELISA. Neat scFv-containing *E. coli* supernatant was applied to each ELISA well. Each glycoprotein was titrated using six 1 in 2 dilutions in PBS from a starting concentration of 10 µg/ml.

Unfortunately, when the scFv were expressed as IgG1[101], none bound complete EBOV GP in solid-phase ELISA (Figure 5-23). As a positive control, KZ52 was expressed as IgG1 alongside the test antibodies using the same expression protocol. KZ52 IgG1 bound EBOV GP, suggesting that the IgG1 expression protocol and the ELISA were working correctly.

Some antibodies were expressed in higher quantities than others (see table in Figure 5-23) and 217_A07 could not be expressed at all.

[101] IgG1 antibodies were expressed using the method described in section 2.8 (page 103).



Antibody	Concentration
KZ52	5.87 µg/ml
215_E06	3.47 µg/ml
215_F09	0.31 µg/ml
215_G07	24.34 µg/ml
216_B01	0.61 µg/ml
216_D01	19.18 µg/ml
217_A09	0.11 µg/ml

Figure 5-23: Specificity of EBOV GP-binding scFv expressed as IgG1. Specificity of IgG1 to complete (glycosylated) EBOV GP tested in solid-phase ELISA. Mean OD_{492} (+/- 1 SD) of each antibody tested in duplicate in neat tissue culture supernatant. The concentration of each IgG1 in neat tissue culture supernatant is shown in the table.

5.7 Discussion

5.7.1 B cell Repertoires of Recent Ebola Survivors

The first aim of this chapter was to identify characteristics of the B cell repertoire which might be linked to a successful immune response to Ebola virus disease. This was done by using high-throughput sequencing to compare the properties of the B cell repertoires of three aid workers who had recently recovered from Ebola virus disease with three age- and gender-matched negative controls (Caucasian Aid Worker Survivor dataset). We hoped that the information gleaned from this investigation would help us to understand the characteristics of a successful anti-Ebola antibody response because, although the immune response against Ebola is not fully understood, it is known that a strong antibody response is linked to recovery from Ebola virus disease (Ksiazek et al., 1999, Baize et al., 1999).

We found a few small but significant differences in V(D)J usage and the CDR3 physicochemical properties between the Ebola survivors and their matched controls at both day 0 and day 28. IGHV1 and IGHV3 were significantly less frequent in the Ebola survivors' repertoires than those of the negative controls and these differences were reflected by the frequency of IGHV1-2, IGHV1-69 and IGHV3-23. IGHV4 was significantly more common in the Ebola survivors' repertoires, and this appeared to be driven by the very large difference in IGHV4-39 usage. IGHV4-39 was approximately three times more common in the B cell repertoires of the Ebola survivors than the negative controls (Figure 5-1, page 218).

There were fewer differences in the light chain repertoire between the sample groups. IGKJ1 was significantly more common in the Ebola survivors than the YF vaccine day 28 controls and IGLJ1 was significantly more common in the Ebola survivors compared to both YF vaccine control time points, whereas IGLJ3 was less common (Figure 5-2, page 219). In the V gene repertoires, IGKV1-39, IGKV4-1 and IGLV2-23 were significantly more frequent in the Ebola survivors' repertoires than the negative controls. As discussed in the previous chapter in section 4.7.1 (page 208), IGKV4-1 is over-represented in autoimmune diseases such as SLE (Dörner et al., 1998) and coeliac disease (Steinsbø et al., 2014).

It was slightly surprising that there were not larger differences between the Ebola survivors and YF vaccine controls (we found no significant differences in heavy and light chain CDR3 properties between the sample groups). However, it has been found previously that Ebola virus disease does not lead to very large clonal expansions and it is thought that the virus may elicit a more polyclonal response than other viral infections such as HIV and influenza (Bornholdt et al., 2016b). Consistent with this theory, we found no clear difference in the

distribution of clonal expansions in the Ebola survivors compared to the YF vaccine controls (Figure 5-3, page 221).

An experimental reason which may partially explain why we found few differences between the Ebola survivors and the negative controls was that, for safety reasons, blood was taken after the donors had cleared the infection. This meant that the libraries were built from donors no longer fighting infection and so the B cell repertoires may have been returning to baseline by the time blood was taken. However, the immune response against Ebola is still detectable several months after the infection has cleared (McElroy et al., 2015, Bornholdt et al., 2016b, Misasi et al., 2016) and in some patients, even three months after primary infection, 3% of circulating IgG+ B cells are still GP-specific (Bornholdt et al., 2016b), therefore we believe that the differences we observed could have been a result of the recent Ebola infection, although differences may have been clearer had we been able to sort the B cells into relevant populations before building the libraries (whole blood was stored in TRIzol which meant that cell sorting was not possible).

Although the differences in V(D)J usage may have been a result of Ebola virus disease, we may just have been measuring inter-individual variation (Galson et al., 2015). When constructing the Caucasian Aid Worker Survivor HTS dataset, it was impossible to obtain blood samples from the Ebola survivors from before they contracted the disease. Therefore, the best negative controls which could be achieved were age- and gender-matched healthy individuals. In order to be more certain of the differences we saw, much larger age- and gender-matched studies would need to take place.

5.7.2 Cross-Class Clones in Ebola Survivors

Despite the experimental limitations discussed above, we did find a very significant difference in the frequency of clonal expansions comprising more than one heavy chain antibody class. We called these clones “cross-class clones” (as opposed to “single-class clones”).

Figure 5-4A (page 223) showed that in the Caucasian Aid Worker Survivor dataset, the frequency of cross-class clones as a proportion of all expanded clones was significantly higher (approximately 10-20%) in all of the Ebola survivors compared to the YF vaccine controls both at day 0 and day 28 (approximately 1-5%). There was no difference in the frequency of cross-class clones between day 0 and day 28 of the YF vaccine controls, indicating that the YF vaccine (our negative control viral challenge) does not induce an increase in the frequency of cross-class clones in the B cell repertoire.

As discussed in the previous section, it must be borne in mind that any differences we see between donors could be inter-individual variation rather than a specific response to Ebola virus disease. However, almost exactly the same frequency of cross-class clones was also observed in a second HTS dataset we built: the African Survivors dataset. Figure 5-9A (page 228) shows that approximately 8-18% of expanded clones in the five African Ebola survivors were cross-class clones. The fact that similarly high frequencies of cross-class clones were observed in two groups of Ebola survivors from different backgrounds strengthens the case that this could be a feature of Ebola virus disease and not merely inter-individual variation.

Interestingly, the high frequency of cross-class clones was maintained in the African survivors across the two time points at which blood was collected (Figure 5-9A) even though the time interval was between three and eight months (Figure 2-4, page 84). It is known that Ebola can hide in immune privileged sites and viral RNA can be present in bodily fluids such as sweat, urine, breast milk and conjunctival fluid (Chughtai et al., 2016) and has been detected in semen up to 15 months after patients leave hospital (Deen et al., 2017). Therefore it is conceivable that, if a high frequency of cross-class clones is indicative of Ebola virus disease, maintenance of a high frequency of cross-class clones might be indicative of the infection having not completely cleared. However, it is unknown whether any of the donors in this study did maintain EBOV in immune privileged sites or if they suffered any viral re-activation. To investigate further it would be interesting to follow survivors over many years.

We have only found one instance in the literature of the quantification of the frequency of cross-class clones using high-throughput sequencing (Kitaura et al., 2017). In this publication, the frequency of clonal expansions comprising more than one heavy chain class were reported to be approximately 1% of expanded clones. This approximately agrees with the frequency we observed in the YF vaccine controls (1-5%), indicating that this is an accurate estimate of the frequency of cross-class clones in healthy individuals.

In both the Aid worker Ebola survivors and the African Ebola survivors, the mean size of cross-class clones was higher than single-class clones (Figure 5-5 & Figure 5-10, pages 224 & 229), however they do not have greater levels of somatic hypermutation (SHM) than IgG single-class clones (Figure 5-7 & Figure 5-11, pages 225 & 229). Sequential class switching to more downstream heavy chain classes can be associated with higher levels of SHM due to increased exposure to AID (Jackson et al., 2014).

One potential clue for the role of cross-class clones comes from the study of the IgE response. In order to create very high affinity IgE, sequential class switching is required via IgG1 (Xiong et al., 2012). Furthermore, IgG1 memory B cells can produce IgE-expressing B cells in the

secondary antibody response (He et al., 2017). We are not aware that this phenomenon has yet been reported for other antibody classes, however it is possible that the use of multiple antibody classes to produce a strong memory response is not just used in the IgE response, but is common to lots of antibody responses. Perhaps our cross-class clones are evidence of this.

Much more experimentation is required to determine if a high frequency of cross-class clones is caused by Ebola virus disease and, if it is, whether a high frequency of cross-class clones is an advantageous response. For stronger evidence that cross-class clones increase in frequency as a result of Ebola infection, the B cell repertoire or more survivors needs to be sequenced. To investigate whether cross-class clones are associated with survival, the B cell repertoires of non-survivors could be sequenced to determine whether a high frequency of cross-class clones is also seen in this patient cohort.

Aside from sequencing more donors, a further experiment to investigate the nature of cross-class clones may be to clone some of the antibodies from cross-class clones and test their specificity to determine if they are Ebola-specific.

5.7.3 Ribosome Display Panning Selection against EBOV GP

The second aim of this chapter was to use ribosome display to identify new antibodies (scFv) which cross-react with glycoprotein (GP) from multiple *Filoviridae* species. Cross-reactive anti-Ebola antibodies are rare and their discovery is important because they could be useful in future therapeutic monoclonal antibody cocktails.

Membrane-bound GP coats the Ebola virus and is used for viral entry into host cells (Baseler et al., 2017); it is also the most common target for therapeutic antibodies. It has been shown previously that cross-reactive anti-GP antibodies can be isolated from human donors who have only been infected with a single species of Ebola virus (Flyak et al., 2016). Therefore, we aimed to isolate new cross-reactive anti-GP antibodies using ribosome display. The ribosome display libraries were built from total RNA isolated from the three Ebola survivors from the Caucasian Aid Worker Survivor dataset.

After a single round of panning selection against EBOV GP, 21 unique EBOV GP-binding scFv were isolated (Figure 5-15, page 236). Of these, seven appeared in the HTS dataset (Table 5-3, page 237). Supporting the hypothesis discussed in the previous section that cross-class clones may be important in the anti-Ebola response, all seven of these scFv were either part of a cross-class clone in the HTS dataset or were of a different class to the one that appeared in the HTS dataset.

The V(D)J gene usage in the selected scFv were relatively limited. Notably, two of the scFv used IGHV4-34. IGHV4-34 is heavily associated with autoimmune diseases such as SLE (Tipton et al., 2015), MS (Baranzini et al., 1999) and RA (Doorenspleet et al., 2014) this may be significant as Ebola virus disease has been linked to autoreactive antibodies, specifically antibodies against dsDNA and heat shock protein 60 (HSP60) (Fausther-Bovendo et al., 2017). It has also been reported that, like autoreactive antibodies, anti-GP antibodies tend to have longer CDR-H3. However, the mean CDR-H3 length of the 17 GP binders we isolated was 14.94 amino acids, which is slightly lower than the mean for these Ebola survivors in the HTS dataset as a whole (15.03 amino acids). Despite this apparent link between IGHV4-34 and EBOV GP specificity, IGHV4-34 was not more frequent in the HTS dataset of the Ebola survivors compared to the YF vaccine controls (Figure 5-1, page 218). But IGHV4-34 did appear to be slightly positively selected in the ribosome display library after the panning selection (Figure 5-18A, page 240).

IGKV5-2 was used in six of the 21 scFv (Table 5-3), however it is rare in the HTS dataset (Figure 5-2, page 219). This suggests that IGKV5-2 may be linked to EBOV GP specificity. However, IGKV5-2 is more common in the PCR3 input library than the HTS library, indicating that primer bias may have led to its over-representation and thus its prevalence in the selected scFv. IGKV5-2 was not more common in the PCR4 output library compared to the PCR3 input library (Figure 5-18B, page 240), indicating that this gene was not positively selected.

None of the genes used in the scFv in Table 5-3 were a lot more common in the PCR4 output dataset (Figure 5-18) and there was no clear directionality observed in the selection (Figure 5-17, page 239) although diversity was reduced in the PCR4 output (Figure 5-16, page 238). However, one round of panning selection may not have been enough to see large trends in selection. Additionally, the anti-GP repertoire has previously been reported to be very diverse (Bornholdt et al., 2016b), therefore large biases and may not be present.

The majority of scFv isolated from the ribosome display were IgA or IgM. It has been reported that IgA and IgM may be important in the response against Ebola (Khurana et al., 2016), however, the frequency of IgA and IgM antibodies was higher in the PCR3 input ribosome display library than the HTS library. This was probably due to primer bias. This may therefore be the reason for the high frequency of IgA and IgM scFv rather than an increased IgA and IgM response in the survivors.

It was concerning that four of the 21 scFv isolated because they bind EBOV GP had sequences which were non-functional either because they are out of frame or incomplete (Table 5-3). 216_A01, 216_A07, 217_A06 and 217_A12 should not have shown binding to EBOV GP and they should not have shown luminescence because stop codons separated the scFv sequence

from the NanoLuc luciferase gene in the pESL plasmid. The quality control checks done to confirm that scFv show specific binding and not every scFv will bind EBOV GP (Figure 5-14, page 234) gave us confidence that the ribosome display system was robust. However we could not find an explanation for why these scFv apparently showed binding and luminescence in the titration ELISAs (Figure 5-15, page 236). However, their luminescence was lower than most of the other scFv.

5.7.4 EBOV GP-Binding scFv may Bind N-linked Glycans

A selection of the EBOV GP-binding scFv were tested against GP from Marburg virus Ravn strain (RAVV) to determine if they are cross-reactive. Marburg virus was chosen as there is less sequence similarity with EBOV than another Ebola virus (Keck et al., 2016) and cross-reactive antibodies which bind Ebola viruses and Marburg viruses are particularly rare. In addition, as a negative control, the scFv were tested against gp120 from HIV. It was expected that if the scFv were monoreactive, they would bind EBOV GP only, if they were cross-reactive they may bind EBOV GP and RAVV GP. It was not expected that they would bind gp120 from HIV.

Very surprisingly, all of the seven scFv bound EBOV GP, RAVV GP and HIV gp120 in the solid-phase ELISA (Figure 5-19, page 241). This suggested that the scFv may have been binding something that all three glycoproteins have in common. EBOV GP and RAVV GP contain both N-linked and O-linked glycans (Lee & Saphire, 2009, Feldmann et al., 1994), however HIV gp120 contains almost exclusively N-linked glycans (Stansell et al., 2015). Furthermore, all three proteins were His-tagged. Therefore, we suspected that the scFv were binding either the N-linked glycans or the His-tags.

To investigate whether the scFv were binding the N-linked glycans, the glycoproteins were deglycosylated using Endo H and PNGase F (Figure 5-21, page 243). Of the seven scFv tested, 215_F09 and 216_B01 showed substantially less binding to the deglycosylated glycoproteins (Figure 5-22, page 244). The other five scFv showed little or no clear difference in binding between the complete and deglycosylated glycoproteins. This may indicate that these scFv are not glycan-specific and may be binding another epitope shared by all three glycoproteins (e.g. the His-tag). However, the luminescence signal from these scFv was low and it may be that the level of scFv expression was too low for a clear result.

Unfortunately, none of these scFv showed binding to complete EBOV GP when expressed by mammalian cells as IgG1 (Figure 5-23, page 245). It is known that variable regions which show binding as scFv do not always maintain their specificity when expressed as IgG (Steinwand et al., 2014). This may be because of differences in structure; the heavy and light chain variable

regions of scFv are joined with a flexible linker peptide instead of the constant region of an antibody. It may also be because of the difference in expression system; scFv are expressed by *E. coli* and are not subject to post-translational modifications, whereas IgG1 expressed by mammalian cells are. Furthermore, the original antibodies of these scFv were IgA and IgM, whereas when they were tested in ELISA they were expressed as IgG1. The constant region can affect the binding affinity of the variable region (Janda et al., 2016) so they may bind if they were expressed as IgA or IgM. The majority (58%) of known anti-glycan antibodies are IgM (Sterner et al., 2016).

Further experiments are needed to determine for certain whether 215_F09 and 216_B01 do bind N-linked glycans, whether they bind them with an affinity that is biologically meaningful, and whether they could be therapeutically useful. However, even if not useful for neutralising filoviruses or HIV, if the antibodies are specific for N-linked glycans, they still may be useful for other purposes as the range of anti-carbohydrate antibodies is very limited, particularly anti-N-linked glycan antibodies. This may be because N-linked glycans are common in mammals and so have low immunogenicity. Furthermore, many N-linked glycans have similar structure, making antibody specificity a problem; many antibodies recognise groups of N-linked glycans rather than individual ones (Sterner et al., 2016).

Due to time constraints, no further experiments were possible. However, we think the following experiments are necessary to clarify the binding specificity of the scFv, whether they can bind as antibodies and whether they may be therapeutically useful:

1. Competition ELISA in the presence of sugars to test if they block scFv binding to the glycoproteins
2. Competition ELISA in the presence of another His-tagged protein to determine whether the scFv bind His-tags
3. Test binding to glycoproteins deglycosylated individually with Endo H or PNGase F to test whether the scFv bind complex glycans (only PNGase F cleaves complex glycans)
4. Use a glycan array to determine the type of glycan the scFv are binding
5. Purify and quantify the scFv in order to compare binding efficacies
6. Determine whether the strength of binding is biologically meaningful (e.g. Biacore)
7. Express the scFv as the original antibody class (IgA or IgM) to determine whether they bind the glycoproteins in this format
8. Perform neutralisation assays using purified scFv or antibody to investigate whether they can neutralise EBOV, RAVV and/or HIV *in vitro*

Chapter 6

Final Discussion and Future Work

Chapter 6: Final Discussion and Future Work

This thesis comprises three chapters all linked by the common theme of antibodies, but each addressing different issues.

6.1.1 Tolerance, Autoreactivity and Polyreactivity

The first aim of this chapter was to identify antibody heavy chain variable region characteristics which are associated with autoreactivity and polyreactivity and could potentially be used as predictors of these phenotypes. A high-throughput sequencing approach was used to identify variable region properties which may be associated with autoreactivity/polyreactivity because they decreased in prevalence as B cells matured from pre-B to naïve. This hypothesis was based on the finding that autoreactive/polyreactive antibodies are selected out of the repertoire at central and peripheral tolerance checkpoints, which are thought to act on the immature and transitional B cell repertoires respectively (Wardemann et al., 2003).

However, the hypothesis that B cells mature linearly from pre B >> immature >> transitional >> naïve and therefore autoreactive/polyreactive characteristics would become sequentially less common in each repertoire was found not necessarily to be the case. We discovered that the B cell variable region properties did not always fit the trajectory of gene usage loss/gain (Figure 3-1, page 140) or CDR-H3 property increase/decrease (Figure 3-2, page 142) which would be expected if the whole B cell population progresses linearly from pre-B to naïve. Our findings suggested that transitional cells may mature into several distinct B cell populations, each with different mean variable region properties and differing roles in the immune system (Martin et al., 2016).

Our data analysis showed that the frequency of IGHV3, IGHD2 and IGHJ6 family usage (Figure 3-1) and also long CDR-H3, hydrophobic CDR-H3 and CDR-H3 with a low Boman index (Figure 3-2A,E&C) decreased as B cells matured, suggesting that heavy chains with these properties may be more likely than average to be autoreactive. Furthermore, CDR-H3 with isoelectric points (pI) between 5.0-6.0 and 7.0-8.5 were very rarely found in any of the B cell repertoires (Figure 3-2D), suggesting that there could be something disadvantageous about these properties; potentially autoreactivity. These hypotheses were tested using immunocytochemistry against HEp-2 cells and it was found that long, hydrophobic CDR-H3, and also aromatic CDR-H3, tended to be associated with cytoplasmic binding (section 3.4, page 153). Furthermore, the underrepresented CDR-H3 pI were associated with nuclear binding (section 3.4.5 & 3.4.6, pages 161 & 162). Anti-nuclear antibodies are heavily associated with

systemic lupus erythematosus (SLE) and it has been shown that CDR-H3 pI is linked to DNA binding in mice (Liang et al., 2004). It would therefore be interesting to see if these antibodies are associated with SLE in humans.

A further surprise finding in this chapter was that, when we came to investigate antibody polyreactivity using solid-phase ELISA, the two polyreactive positive controls we used from the literature (Meffre et al., 2004, Wardemann et al., 2003) showed the same binding pattern in solid-phase ELISA even in the absence of antigen on the plate (Figure 3-11, page 164). Furthermore, this apparent polyreactive phenotype disappeared when using certain protein or high-salt blocking buffers (Figure 3-12, page 165). We followed this up with experiments using our own antibodies (GF antibodies) and found that antibodies displaying this unusual polyreactive phenotype in solid-phase ELISA did not appear to bind these antigens in other formats such as soluble ELISA (Figure 3-16, page 168) or immunocytochemistry (Figure 3-17, page 169). In the literature, this blocking buffer-dependent phenotype has been described in various human sera (Fujii et al., 1989, Jørgensen et al., 2005, Terato et al., 2014, Güven et al., 2014) and it is associated with inflammatory mediators (Güven et al., 2014). We decided to refer to antibodies displaying this phenotype as “promiscuous” to distinguish them from antibodies which bind multiple unrelated antigens specifically.

We believe that antibodies which show a polyreactive phenotype when using the protein-free blocking solid-phase ELISA protocol described in Tiller et al. (2008) might in fact be promiscuous. These antibodies are selected against at tolerance checkpoints (Wardemann et al., 2003) and are associated with autoimmune diseases such as RA (Samuels et al., 2005) and SLE (Yurasov et al., 2005, Yurasov et al., 2006). They are also more common in people with conditions such as a deficiency in AID (Meyers et al., 2011, Cantaert et al., 2015) or CD40L (Hervé et al., 2007). If these antibodies are promiscuous, this would suggest that the promiscuous phenotype is very physiologically relevant. However, the nature of their interactions with antigen remains to be fully understood.

From modelling the GF antibodies, promiscuous antibodies seem to have highly structured CDR-H3 (Laffy et al., 2016). Experiments with CT0037, CT0053 and CT0054 showed that the promiscuous phenotype might be linked to aggregation (Figure 3-21, page 173) and all of the promiscuous antibodies we found bound to the cytoplasm of HEp-2 cells, indicating that they may be autoreactive. Further investigation is needed to fully characterise promiscuous antibodies and determine their role in the immune response.

6.1.2 Light Chain Selection and Isotype Properties

The antibody variable region is composed of a heavy and a light chain. However, the contribution of the light chain is sometimes overlooked as the heavy chain (particularly the highly diverse CDR-H3) often makes most of the contacts with antigen. However, light chains do contribute important contacts (Kunik & Ofran, 2013) and are central to receptor editing (Nemazee, 2006). Furthermore, some light chain properties are associated with autoimmune disease (Dörner et al., 1998, Steinsbø et al., 2014, Samuels et al., 2005).

The first aim of this chapter was to investigate how the properties of light chain variable regions change as B cells mature to determine whether light chains may be subject to the same tolerance selection pressures as heavy chains. We found very little evidence of positive or negative selection of light chain variable region properties as B cells matured (section 4.2, page 192), suggesting that the selection pressures may be different on light chains than on heavy chains. There was slightly more evidence of selection in the lambda repertoire than the kappa repertoire. This made sense as lambda light chains tend to be rearrange after kappa light chains during receptor editing and lambda light chains are better than kappa light chains at rescuing autoreactive heavy chains (Wardemann et al., 2004).

The fact that there are two classes of light chain (kappa and lambda) is often overlooked in the literature. The different roles of the heavy chain classes have been studied extensively, but light chains are sometimes considered as one group with no distinction made between the classes. However, there is evidence to suggest that kappa and lambda light chains may play different roles in the immune system. For example, kappa light chains are more common than lambda in the peripheral blood (Mole et al., 1994, Brauninger et al., 2001, Montano & Morrison, 2002), however lambda light chains are more frequently found in secreted antibodies (Mole et al., 1994). There are also biases observed in certain antigen-selected populations, such as anti-HIV Env antibodies (Sajadi et al., 2016).

The second aim of this chapter was to compare the physicochemical properties of kappa and lambda CDR-L3 and we found that kappa and lambda light chains have very different CDR-L3 physicochemical properties. For example, lambda CDR-L3 are significantly longer and more hydrophobic than kappa CDR-L3 (Figure 4-4, page 199). Kappa and lambda CDRs also have different structural compositions (Figure 4-9A, page 206). This difference suggests that kappa and lambda light chains might have different roles in the immune system and that they might have different heavy chain pairing preferences, perhaps to complement the contrasting physicochemical properties of the kappa and lambda light chains.

Whether heavy-light chain pairing is random is debated in the literature (DeKosky et al., 2016, Glanville et al., 2009, Brezinschek et al., 1998, Jayaram et al., 2012), however patterns in heavy-light chain pairing have not yet been identified. Therefore we compared the VDJ usage and CDR-H3 properties of heavy chains paired with kappa and lambda light chains to determine if there were any biases which might suggest that certain heavy chains paired well with kappa or lambda light chains. However, we found no difference in the properties of the heavy chains (Figure 4-7, page 204) and this was supported by structural data (Figure 4-9B, page 206). This supports the hypothesis that heavy-light chain pairing is random. However, much more work is needed to confirm whether heavy-light chain pairing truly is random, for example other properties or parts of the variable regions may drive heavy-light chain pairing.

The final finding of this chapter was that there is a strong positive correlation between the number of N-nucleotides in kappa, lambda and heavy chain CDR3 regions within individuals (Figure 4-10, page 207). Furthermore, the difference in mean N-nucleotide addition between some of the donors was as much as 2.5 nt, 4.3 nt and 3.0 nt at the kappa, lambda and heavy chain CDR3 regions respectively; enough to encode an additional amino acid. As long CDR-H3 with lots of N-nucleotide addition is associated with autoreactivity, it would be very interesting to examine the mean N-nucleotide addition in various patient groups with different autoimmune diseases to determine whether people with a higher mean N-nucleotide addition are more prone to autoimmune disease. It would also be interesting to investigate if differing amounts of N-nucleotide addition led to differing responses to infection, for example the likelihood of developing broadly neutralising antibodies to HIV (Pancera et al., 2010).

6.1.3 The Antibody Response against Ebola

The immune response to Ebola virus disease is not completely understood, however serology has shown that a strong antibody response is associated with survival (Ksiazek et al., 1999, Baize et al., 1999). Therefore the first aim of this chapter was to learn more about the B cell response to the most deadly strain of Ebola virus, Zaire (abbreviated to EBOV), by studying the properties of the B cell repertoire in recent EBOV survivors and age- and gender-matched negative controls who had received a yellow fever (YF) vaccine.

Only small differences were found in the V(D)J usage between the recent Ebola survivors and YF controls (section 5.2.1, page 217) and no significant differences were found in CDR3 properties (section 5.2.2, page 219). This may have been because the anti-Ebola response is quite polyclonal (Bornholdt et al., 2016b). Alternatively, it may have been a consequence of only being able to use blood samples taken after the donors had recovered or, because it was

not possible to obtain negative control blood samples from the Ebola survivors, inter-individual variation was a factor (Galson et al., 2015).

However, one significant finding of this data analysis was that the frequency of clonal expansions comprising multiple heavy chain classes ("cross-class clones") was significantly higher in the Ebola survivors than the YF negative controls (Figure 5-4, page 223) or figures stated in the literature (Kitaura et al., 2017). Furthermore, a similarly high frequency of cross-class clones was observed in the African Ebola survivors (Figure 5-9, page 228). The fact that very similar levels of cross-class clones were found in completely separate patient groups suggests that this is not simply inter-individual variation and it could be that cross-class clones are associated with Ebola virus disease. More samples and further experiments are required to confirm this link, and to determine whether a high frequency of cross-class clones is advantageous in the immune response against Ebola virus disease.

The second aim of this chapter was to use ribosome display to identify new anti-Ebola GP scFv and also cross-reactive scFv. We identified 21 unique EBOV GP-binding scFv (Figure 5-15, page 236). There was little evidence of positive selection of the V(D)J genes used in the scFv (Figure 5-18, page 240), however seven of the scFv did occur in the HTS dataset. Most were part of expanded clones and all appeared to be part of cross-class clones (Table 5-3, page 237). Surprisingly, not only did these seven scFv bind EBOV GP, they also showed binding to RAVV GP and HIV gp120 (Figure 5-19, page 241). Furthermore, two of these scFv lost binding to the deglycosylated forms of these glycoproteins (Figure 5-22, page 244) indicating that they may be binding the N-linked glycans. However, unfortunately none of these variable regions bound complete EBOV GP when expressed as IgG1 (Figure 5-23, page 245).

If these scFv genuinely are cross-reactive, it would certainly be worth investigating them further with a view to any therapeutic potential they may have. This is because a large drawback of the anti-Ebola monoclonal antibody cocktails currently in use is that they are only specific to EBOV. EBOV is the most lethal Ebola virus, however other, lesser studied species also pose a substantial threat (Burk et al., 2016) and it is important to develop treatments which could be used in any *Ebolavirus*, or ideally *Filoviridae*, outbreak. Furthermore, if the antibodies are specific to an N-linked glycan this would also be very useful as antibodies which bind N-linked glycans are rare.

This was the final chapter of work and, due to time constraints, we were unable to conduct the experiments required to determine whether the scFv are truly cross-reactive, whether they bind N-linked glycans, whether they bind when expressed as IgA or IgM and whether they may be therapeutically useful (the list of immediate experiments required is given on page 252).

6.1.4 Concluding Remarks

The work in this thesis has examined the properties of both antigen-naïve and antigen-experienced B cell repertoires in a number of different contexts: the antigen-naïve repertoires of B cell development, the light chain repertoires separated by kappa and lambda class and the antigen-experienced repertoires of recent Ebola survivors. In doing so, we have identified some antibody characteristics which may be associated with autoreactivity, investigated promiscuous antibodies, highlighted the very different properties of kappa and lambda light chains, investigated heavy-light chain pairing, identified a high frequency of cross-class clones in Ebola survivors and have identified some potentially cross-reactive or N-linked glycan-specific scFv.

As a whole, perhaps the most striking inference that can be drawn from this work is that antibody variable regions are so diverse and complex that it is currently not possible to state any hard and fast rules governing antibody specificity. Only a few general associations between variable region properties and certain binding specificities can be identified. Hopefully, as computing power and molecular modelling techniques continue to improve, spatial and environmental factors can be more easily taken into account when considering antibody specificity and we will understand ever more about the subject. Something which will help immensely with the future investigation of antibody specificity is the ability to conduct paired heavy-light chain sequencing. As high-throughput sequencing techniques become capable of longer and more precise sequencing reads and with the development of technologies such as single-cell microfluidics, paired sequencing will hopefully become the norm in the near future. This type of sequencing will also be extremely helpful in settling whether heavy-light chain pairing is truly random and for understanding more about the role of light chains in the antibody response.

A lot of questions have been raised in this thesis regarding, among other things, the nature of polyreactivity, the role of kappa and lambda light chains in the antibody repertoire and the purpose of cross-class clones in the repertoires of Ebola survivors. We hope that this work paves the way for further investigation into these topics, improving our understanding of both antigen-naïve and antigen-experienced human B cell repertoires and hopefully enhancing our ability to treat and prevent disease.

Appendix

Appendix

A.1 Primer List

A.1.1 High-Throughput Sequencing Primers: B cell Development Dataset

Table A-1: Primers used for B cell Development dataset HTS PCR1.

Primer name	Sequence (5' to 3')	Direction	
IGHV1	CCTCAGTGAAGGTCTCCTGCAAGG	Forward	IgM primers
IGHV2	TCCTGCGCTGGTGAAACCCACACA	Forward	
IGHV3	GGTCCCTGAGACTCTCCTGTGCA	Forward	
IGHV4	TCGGAGACCCTGTCCCTCACCTGC	Forward	
IGHV5	CAGTCTGGAGCAGAGGTGAAA	Forward	
IGHV6	CCTGTGCCATCTCCGGGGACAGTG	Forward	
IGHM	GGGGAATTCTCACAGGAGAC	Reverse	
IGKV1	CATCCAGWTGACCCAGTCTCC	Forward	IgK primers
IGKV2	GATATTGTGATGACCCAGWCT	Forward	
IGKV3	GACRCAGTCTCCAGCCACCCTG	Forward	
IGKV4	GACATCGTGATGACCCAGTCT	Forward	
IGKV5	GAAACGACACTCACGCAGTCT	Forward	
IGKV6	GAAATTGTGCTGACTCAGTCT	Forward	
IGKC	CCTTCCACTGTACTTTGGCCTC	Reverse	
IGLV1	CAGTCTGTGCTGACKCAGCC	Forward	IgL primers
IGLV2	CAGTCTGCCCTGACTCAGCC	Forward	
IGLV3	CCTATGAGCTGACWCAGCCAC	Forward	
IGLV4/5	CAGCCTGTGCTGACTCARYC	Forward	
IGLV6	CCAGNCTGTGSTGACTCAG	Forward	
IGLC	GCCACTGTCACRGCTCCCGGG	Reverse	

Table A-2: Primers used for B cell Development dataset HTS PCR2. MID1 is used as an example. MID sequences are shown in blue lowercase and the full list of 12 MIDs is shown in Table A-3.

Primer name	Sequence (5' to 3')	Direction
MID1_IGHV1	acgagtgcgtCCTCAGTGAAGGTCTCCTGCAAGG	Forward
MID1_IGHV2	acgagtgcgtTCCTGCGCTGGTGAAACCCACACA	Forward
MID1_IGHV3	acgagtgcgtGGTCCCTGAGACTCTCCTGTGCA	Forward
MID1_IGHV4	acgagtgcgtTCGGAGACCCTGTCCCTCACCTGC	Forward
MID1_IGHV5	acgagtgcgtCAGTCTGGAGCAGAGGTGAAA	Forward
MID1_IGHV6	acgagtgcgtCCTGTGCCATCTCCGGGGACAGTG	Forward
MID1_IGHM	acgagtgcgtCAGGAGACGAGGGGGAAAAGG	Reverse
MID1_IGKV1	acgagtgcgtCATCCAGWTGACCCAGTCTCC	Forward
MID1_IGKV2	acgagtgcgtGATATTGTGATGACCCAGWCT	Forward
MID1_IGKV3	acgagtgcgtGACRCAGTCTCCAGCCACCCTG	Forward
MID1_IGKV4	acgagtgcgtGACATCGTGATGACCCAGTCT	Forward
MID1_IGKV5	acgagtgcgtGAAACGACACTCACGCAGTCT	Forward
MID1_IGKV6	acgagtgcgtGAAATTGTGCTGACTCAGTCT	Forward
MID1_IGKC	acgagtgcgtCCTTCCACTGTACTTTGGCCTC	Reverse
MID1_IGLV1	acgagtgcgtCAGTCTGTGCTGACKCAGCC	Forward
MID1_IGLV2	acgagtgcgtCAGTCTGCCCTGACTCAGCC	Forward
MID1_IGLV3	acgagtgcgtCCTATGAGCTGACWCAGCCAC	Forward
MID1_IGLV4/5	acgagtgcgtCAGCCTGTGCTGACTCARYC	Forward
MID1_IGLV6	acgagtgcgtCCAGNCTGTGSTGACTCAG	Forward
MID1_IGLC	acgagtgcgtGCCACTGTCACRGCTCCCGGG	Reverse

Table A-3: MID sequences used on PCR2 primers for B cell Development dataset HTS. PCR2 primers using MID1 are shown in Table A-2.

MID	Sequence (5' to 3')
MID1	acgagtgcgt
MID2	acgctcgaca
MID3	agacgcactc
MID4	agcactgtag
MID5	atcagacacg
MID6	atatcgcgag
MID7	cgtgtctcta
MID8	ctcgcgtgtc
MID9	tagtatcagc
MID10	tctctatgcg
MID11	tgatacgtct
MID12	tactgagcta

A.1.2 High-Throughput Sequencing Primers: Ebola Response Dataset

Table A-4: SmartNNN primer used for reverse transcription in Ebola HTS. N represents any nucleotide

Primer Name	Sequence (5' to 3')	Direction
SmartNNN	AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNUNNNNUCTTGGGGG	Forward

Table A-5: Primers used in PCR1 for Ebola HTS.

Primer Name	Sequence (5' to 3')	Direction
Smart20	CACTCTATCCGACAAGCAGTGGTATCAACGCAG	Forward
IGHA-1st	CTTCACGTGGCATGTCACGGAC	Reverse
IGHG-1st	TCCTGAGGACTGTAGGACAGC	Reverse
IGHM-1st	TGACGTCCTTGGAAGGCAGCAG	Reverse
IGKC-1st	GTAGTCTGCTTTGCTCAGCGTCAG	Reverse
IGLC-1st	ACGGTGCTCCCTTCATGCGTG	Reverse

Table A-6: PID-labelled primers used in PCR2 for Ebola HTS. PID1 is used as an example and is shown in blue lowercase. All PIDs used are listed in Table A-7.

Primer Name	Sequence (5' to 3')	Direction
PID1_Step_1	ggtagtcatgagtcgacactaCACTCTATCCGACAAGCAGT	Forward
PID1_IGHA-2nd	ccatcgcgatctatgcacacgTACAGGTCCCCGGAGGCATC	Reverse
PID1_IGHG-2nd	ccatcgcgatctatgcacacgGACAGCYGGGAAGGTGTGCAC	Reverse
PID1_IGHM-2nd	ccatcgcgatctatgcacacgCGGGTACTGCTGATGTCAGAG	Reverse
PID1_IGKC-2nd	ccatcgcgatctatgcacacgTGAGGCTGTAGGTGCTGTCCTTG	Reverse
PID1_IGLC-D-2nd	ccatcgcgatctatgcacacgTTCATGCGTGACCCGGCAGC	Reverse

Table A-7: PIDs used on PCR2 primers to label Ebola HTS dataset samples. PID1-labelled PCR2 primers are shown in Table A-6.

PID	Forward (5' to 3')	Reverse (5' to 3')
PID1	ggtagtcatgagtcgacacta	ccatcgcgatctatgcacacg
PID2	ggtagtatctatcgtatacgc	ccatctgcagtcgagatacat
PID4	ggtagacgtacgctcgtcata	ccatctacagcgacgtcatcg
PID9	ggtagtcatgcacgtctcgct	ccatcagtatcacagtcgctg
PID17	ggtagcacgtcactagagcga	ccatccagacgtgactgatat
PID22	ggtaggtgctgagcatcagac	ccatctgagacatactgagtg
PID23	ggtagcactgatcgatatgca	ccatcatgtgcactagtgtac
PID33	ggtagatacagcacagatgtg	ccatcgagtcgtatcgctcat
PID44	ggtagctcgatacgtgtagct	ccatctgtcagtagatgactc
PID45	ggtaggtgtctagacagctgt	ccatctcgtacgagatcgaca

A.1.3 High-Throughput Sequencing Primers: Ribosome Display Libraries

Table A-8: PID-labelled primers used to amplify ribosome display inputs and outputs. PIDs are shown in blue and lowercase and are the same as the ones shown in Table A-7.

Primer Name	Sequence (5’ to 3’)
RDT7AB_PID1	ggtagtcatgagtcgacactaGCTAATACGACTCACTATAGGGAGTCG
NotK_PID1	ccatcgcgatctatgcacacgAGATTTCAACTGCTCATCAGATGG
NotL_PID1	ccatcgcgatctatgcacacgTTGGAGCTCCTCAGAGGAG
RDT7AB_PID4	ggtagacgtacgctcgtcataGCTAATACGACTCACTATAGGGAGTCG
NotK_PID4	ccatctacagcgacgtcatcgAGATTTCAACTGCTCATCAGATGG
NotL_PID4	ccatctacagcgacgtcatcgTTGGAGCTCCTCAGAGGAG
RDT7AB_PID9	ggtagtcatgcacgtctcgctGCTAATACGACTCACTATAGGGAGTCG
NotK_PID9	ccatcagtatcacagtcgctgAGATTTCAACTGCTCATCAGATGG
NotL_PID9	ccatcagtatcacagtcgctgTTGGAGCTCCTCAGAGGAG
RDT7AB_PID22	ggtaggtgctgagcatcagacGCTAATACGACTCACTATAGGGAGTCG
NotK_PID22	ccatctgagacatactgagtgAGATTTCAACTGCTCATCAGATGG
NotL_PID22	ccatctgagacatactgagtgTTGGAGCTCCTCAGAGGAG

A.1.4 Ribosome Display Primers

Table A-9: Primers used for ribosome display PCR1. Degenerate nucleotides[102] are in bold and underlined. Gene-specific sequences are in black, overhang sequences are in red.

Name	Primer sequence (5' to 3')	Ta (°C)	Direction
VH1	CAGGTCCAGCT**K**GT**R**CAGTCTGG	70	Forward
VH1257	CAGGTGCAGCTGGTG**S**A**R**TCTGG	72	Forward
VH2	CAG**R**TCACCTTGAAGGAGTCTG	66	Forward
VH3A	GAGGTGCAGCTG**K**TGGAG**W**CC	70	Forward
VH3B	GAGGTGCAGCTG**K**TGGAG**W**CT	70	Forward
VH4	CAGGTGCAGCTGCAGGAGTC**S**G	72	Forward
VH4DP63	CAGGTGCAGCTACAGCAGTGGG	72	Forward
VH6	CAGGTACAGCTGCAGCAGTCA	71	Forward
GCH1link	GGAGGAGGGCGCCAGGGGGAAGACCGATGGGCCC		Reverse
MCH1link	GGAGGAGGGCGCCAGGGGGAAAAGGGTTGGGGCGGATGCACTCCC		Reverse
ACH1link	GGAGGAGGGCGCCAGGGGGAAGACCTTGGGGCTGGTCGGGGATGC		Reverse
VK1	GACATCC**R**GDTGACCCAGTCTCC	67	Forward
VK2346	GATATTGTG**M**TGACBCAG**W**CTCC	60	Forward
VK36	GAAATTGT**RW**TGAC**R**CAGTCTCC	59	Forward
VK5	GAAACGACACTCACGCAGTCTC	66	Forward
KC	AGACTCTCCCCTGTTGAAGCTCTTTGTGACGGGCGAGCTCAG		Reverse
VL1	CAGTCTGT**S**BTGACGCAGCCGCC	72	Forward
VL1459	CAGCCTGTGCTGACTCA**RY**C	64	Forward
VL15910	CAGCC**W**GKGCTGACTCAGCC**M**CC	72	Forward
VL2	CAGTCTG**YY**CTGA**Y**TCAGCCT	62	Forward
VL3a	TCCTATG**W**GCTGAC**W**CAGCCAC	67	Forward
VL3b	TCCTATGAGCTGA**YR**CAGC**Y**ACC	63	Forward
VL3DLP16	TCCTCTGAGCTGA**S**TCAGGA**S**CC	70	Forward
VL6	AATTTTATGCTGACTCAGCCCC	63	Forward
VL78	CAGDCTGTGGTGAC**Y**CAGGAGCC	70	Forward
LC	AGAAGATTCTGTAGGGGCCACTGTCTTCTCCACGG		Reverse

[102] Degenerate code: B=C/G/T, D=A/G/T, H=A/C/T, K=G/T, M=A/C, R=A/G, S=G/C, V=A/C/G, W=A/T, Y=C/T

Table A-10: Primers used for ribosome display PCR2. Degenerate nucleotides[103] are in bold and underlined. Gene-specific sequences are in black, overhang sequences are in red.

Name	Primer sequence (5' to 3')	Ta (°C)	Direction
VH1A	AGTCGCCGCCATGGCCCAGGTCCAGCTKGT**R**CAGTCTGG	70	Forward
VH1257A	AGTCGCCGCCATGGCCCAGGTGCAGCTGGTG**S**A**R**TCTGG	72	Forward
VH2A	AGTCGCCGCCATGGCCCAG**R**TCACCTTGAAGGAGTCTG	66	Forward
VH3AA	AGTCGCCGCCATGGCCGAGGTGCAGCTGKTGGAG**W**C	70	Forward
VH4A	AGTCGCCGCCATGGCCCAGGTGCAGCTGCAGGAGTC**S**G	72	Forward
VH4DP63A	AGTCGCCGCCATGGCCCAGGTGCAGCTACAGCAGTGGG	72	Forward
VH6A	AGTCGCCGCCATGGCCCAGGTACAGCTGCAGCAGTCA	71	Forward
GMACH1link	GGAGGAGGGCGCCAGGGGGAA		Reverse
VK1link	CCCCCTGGCGCCCTCCTCCGACATCC**R**GDTGACCCAGTCTCC	67	Forward
VK2346link	CCCCCTGGCGCCCTCCTCCGATATTGTG**M**TGACBCAG**W**CTCC	60	Forward
VK36link	CCCCCTGGCGCCCTCCTCCGAAATTGT**RW**TGAC**R**CAGTCTCC	59	Forward
VK5link	CCCCCTGGCGCCCTCCTCCGAAACGACACTCACGCAGTCTCC	66	Forward
KCNotI	CCCGTTGGCCTTGCTTGCGGCCGCAGACTCTCCCCTGTTGAAG CTCTTTGTGACGGG		Reverse
VL1link	CCCCCTGGCGCCCTCCTCCCAGTCTGT**SB**TGACGCAGCCGCC	72	Forward
VL1459link	CCCCCTGGCGCCCTCCTCCCAGCCTGTGCTGACTCA**RY**C	64	Forward
VL15910link	CCCCCTGGCGCCCTCCTCCCAGCC**W**G**K**GCTGACTCAGCC**M**CC	72	Forward
VL2link	CCCCCTGGCGCCCTCCTCCCAGTCTG**YY**CTGA**Y**TCAGCCT	62	Forward
VL3alink	CCCCCTGGCGCCCTCCTCCTCCTATG**W**GCTGAC**W**CAGCCAC	67	Forward
VL3blink	CCCCCTGGCGCCCTCCTCCTCCTATGAGCTGA**YR**CAGC**Y**ACC	63	Forward
VL3DLP16link	CCCCCTGGCGCCCTCCTCCTCCTCTGAGCTGA**S**TCAGGA**S**CC	70	Forward
VL6link	CCCCCTGGCGCCCTCCTCCAATTTTATGCTGACTCAGCCCC	63	Forward
VL78link	CCCCCTGGCGCCCTCCTCCCAG**D**CTGTGGTGAC**Y**CAGGAGCC	70	Forward
LCNotI	CCCGTTGGCCTTGCTTGCGGCCGCAGAAGATTCTGTAGGGGCC AC		Reverse

[103] Degenerate code: B=C/G/T, D=A/G/T, H=A/C/T, K=G/T, M=A/C, R=A/G, S=G/C, V=A/C/G, W=A/T, Y=C/T

Table A-11: Primers used for ribosome display PCR3. Gene-specific sequences are in black, overhand sequences are in red.

Name	Primer sequence (5' to 3')	Direction
RDT7AB	GCTAATACGACTCACTATAGGGAGTCGCCGCCATGGCCC	Forward
ASKANG	CCCGTTGGCCTTGCTTG	Reverse

Table A-12: Primers used for ribosome display reverse transcription. Degenerate nucleotides[104] are in bold and underlined.

Name	Primer sequence (5' to 3')	Direction
HuκF	TCCAGATTTCAACTGCTCATCAGATGGCGG	Reverse
HuλF	GGCTTGGAGCTCCTCAGAGGAGGG**Y**GGGAA	Reverse

Table A-13: Primers used for ribosome display PCR4. Gene-specific sequences are in black, overhand sequences are in red.

Name	Primer sequence (5' to 3')	Direction
RDT7AB_short	GCTAATACGACTCACTATAGGGAGTCG	Forward
NotK	TATAGCGGCCGCAGATTTCAACTGCTCATCAGATGG	Reverse
NotL	ATATGCGGCCGCTTGGAGCTCCTCAGAGGAG	Reverse

Table A-14: Primers used for ribosome display colony PCR and sequencing. T7F and T7R are used for colony PCR. T7F and pESL-seqR are used for sequencing.

Name	Primer sequence (5' to 3')	Direction
T7F	TAATACGACTCACTATAGGG	Forward
T7R	GCTAGTTATTGCTCAGCGG	Reverse
pESL_seqR	CATGAATATCAATTTTCAGGC	Reverse

[104] Degenerate code: B=C/G/T, D=A/G/T, H=A/C/T, K=G/T, M=A/C, R=A/G, S=G/C, V=A/C/G, W=A/T, Y=C/T

A.1.5 Polymerase Incomplete Primer Extension (PIPE) Cloning Primers

Table A-15: pVITRO1-dV-IgG1 plasmid PIPE linearization primers.

Primer name	Sequence (5' to 3')	Direction
Linear_Hrev	GGAGTGCGCGCCTGTGGCGGCCGCCACCAAGAAGAGGATC	Reverse
Linear_IgG1fwd	GCTAGCACCAAGGGCCCATCGGTCTTCCCCCTGGCACCCT	Forward
Linear_Lrev	AGAGTCGACTCCGGATCCATAAGCAAGGAGTCCGAGGAGA	Reverse
Linear_Lfwd	CTAGGTCAGCCCAAGGCGGCGCCCTCGGTCACTCTGTTCC	Forward
Linear_Krev	ACCGCGGCTAGCTGGAACCCAGAGCAGCAGAAACCCAATG	Reverse
Linear_Kfwd	CGTACGGTGGCGGCGCCATCTGTCTTCATCTTCCCGCCAT	Forward

Table A-16: PIPE cloning primers for variable region amplification. The variable region-complementary regions are in capital letters and the amino acid sequences they encode are indicated in the primer names. The pVITRO1 plasmid-complementary regions are in lowercase.

Primer name	Primer sequence (5' to 3')	Chain	Direction
QVQLV_VH_FW	cacaggcgcgcactccCAGGTGCAGCTGGTG	IgH	Forward
EVQLL_VH_FW	cacaggcgcgcactccGAGGTGCAGCTGTTG	IgH	Forward
QVTLK_VH_FW	cacaggcgcgcactccCAGGTCACCTTGAAG	IgH	Forward
QVQLQ_VH_FW	cacaggcgcgcactccCAGGTGCAGCTGCAG	IgH	Forward
EVQLV_VH_FW	cacaggcgcgcactccGAGGTCCAGCTGGTA	IgH	Forward
VTVSS_IgG1_RV	ggcccttggtgctagcTGAGGAGACGGTGAC	IgH	Reverse
DIQLT_VK_FW	tccagctagccgcggtGACATCCAGTTGACC	IgK	Forward
DIQMT_VK_FW	tccagctagccgcggtGACATCCAGATGACC	IgK	Forward
ETTLT_VK_FW	tccagctagccgcggtGAAACGACACTCACG	IgK	Forward
DIVLT_VK_FW	tccagctagccgcggtGACATCGTGTTGACG	IgK	Forward
EIVLT_VK_FW	tccagctagccgcggtGAAATTGTGTTGACG	IgK	Forward
DIRMT_VK_FW	tccagctagccgcggtGACATCCGGATGACC	IgK	Forward
DIRLT_VK_FW	tccagctagccgcggtGACATCCGGTTGACC	IgK	Forward
ELVMT_VK_FW	tccagctagccgcggtGAGCTCGTGATGACC	IgK	Forward
RLEIK_IgK_RV	gcgccgccaccgtacgTTTAATCTCCAGTCG	IgK	Reverse
KVDIK_IgK_RV	gcgccgccaccgtacgTTTGATATCCACTTT	IgK	Reverse
KLEIK_IgK_RV	gcgccgccaccgtacgTTTGATCTCCAGCTT	IgK	Reverse
KVEIK_IgK_RV	gcgccgccaccgtacgTTTGATTTCCACCTT	IgK	Reverse
SSELS_VL_FW	atccggagtcgactctTCCTCTGAGCTGAGT	IgL	Forward
SYELM_VL_FW	atccggagtcgactctTCCTATGAGCTGATG	IgL	Forward
SSELT_VL_FW	atccggagtcgactctTCTTCTGAGCTGACT	IgL	Forward
QSVLT_VL_FW	atccggagtcgactctCAGTCTGTGCTGACT	IgL	Forward
TKLTV_IgL_RV	ccttgggctgacctagGACGGTCAGCTTGGT	IgL	Reverse
TKVTV_IgL_RV	ccttgggctgacctagGACGGTCAGCTTGGT	IgL	Reverse

Table A-17: PIPE cloning primer combinations for each antibody expressed as IgG1.

	Heavy chain amplification		Light chain amplification	
Antibody	Forward primer	Reverse primer	Forward primer	Reverse primer
215_B12	QVTLK_VH_FW	VTVSS_IgG1_RV	ETTLT_VK_FW	KVEIK_IgK_RV
215_E06	QVQLQ_VH_FW	VTVSS_IgG1_RV	ETTLT_VK_FW	KLEIK_IgK_RV
215_F09	QVTLK_VH_FW	VTVSS_IgG1_RV	ETTLT_VK_FW	KLEIK_IgK_RV
215_F11	QVQLV_VH_FW	VTVSS_IgG1_RV	SSELS_VL_FW	TKLTV_IgL_RV
215_G01	QVQLQ_VH_FW	VTVSS_IgG1_RV	SSELS_VL_FW	TKLTV_IgL_RV
215_G07	QVQLQ_VH_FW	VTVSS_IgG1_RV	SYELM_VL_FW	TKVTV_IgL_RV
215_G10	QVQLV_VH_FW	VTVSS_IgG1_RV	SSELT_VL_FW	TKLTV_IgL_RV
216_A02	QVQLQ_VH_FW	VTVSS_IgG1_RV	DIVLT_VK_FW	KLEIK_IgK_RV
216_A05	EVQLV_VH_FW	VTVSS_IgG1_RV	ETTLT_VK_FW	KVEIK_IgK_RV
216_B01	QVQLQ_VH_FW	VTVSS_IgG1_RV	EIVLT_VK_FW	RLEIK_IgK_RV
216_D01	QVQLV_VH_FW	VTVSS_IgG1_RV	SSELS_VL_FW	TKLTV_IgL_RV
217_A01	EVQLV_VH_FW	VTVSS_IgG1_RV	ETTLT_VK_FW	KLEIK_IgK_RV
217_A07	EVQLV_VH_FW	VTVSS_IgG1_RV	DIRMT_VK_FW	KVEIK_IgK_RV
217_A09	QVQLQ_VH_FW	VTVSS_IgG1_RV	ETTLT_VK_FW	KLEIK_IgK_RV
217_A10	QVQLV_VH_FW	VTVSS_IgG1_RV	DIRLT_VK_FW	KVEIK_IgK_RV
217_B05	EVQLL_VH_FW	VTVSS_IgG1_RV	QSVLT_VL_FW	TKLTV_IgL_RV
GF01	QVQLV_VH_FW	VTVSS_IgG1_RV	DIQLT_VK_FW	RLEIK_IgK_RV
GF05	EVQLL_VH_FW	VTVSS_IgG1_RV	DIQMT_VK_FW	KVDIK_IgK_RV
GF07	QVQLV_VH_FW	VTVSS_IgG1_RV	DIQMT_VK_FW	KLEIK_IgK_RV
GF09	QVQLV_VH_FW	VTVSS_IgG1_RV	DIQMT_VK_FW	RLEIK_IgK_RV
GF16	QVQLV_VH_FW	VTVSS_IgG1_RV	DIQMT_VK_FW	KLEIK_IgK_RV
KZ52	EVQLL_VH_FW	VTVSS_IgG1_RV	ELVMT_VK_FW	KVEIK_IgK_RV

Table A-18: pVITRO1 colony PCR and sequencing primers.

Primer name	Sequence (5' to 3')	Direction
pVITRO1F	TTTTGAGCGGAGCTAATTCTCGGG	Forward
pVITRO1R	AAAAAACCTCCCACACCTCC	Reverse
pVITRO2F	GAGGCTAATTCTCAAGCCTC	Forward
pVITRO2R	TCTAGACCTGGAAAGACCAG	Reverse

A.2 Vector Maps

A.2.1 pVITRO1 IgG1 Expression Plasmids

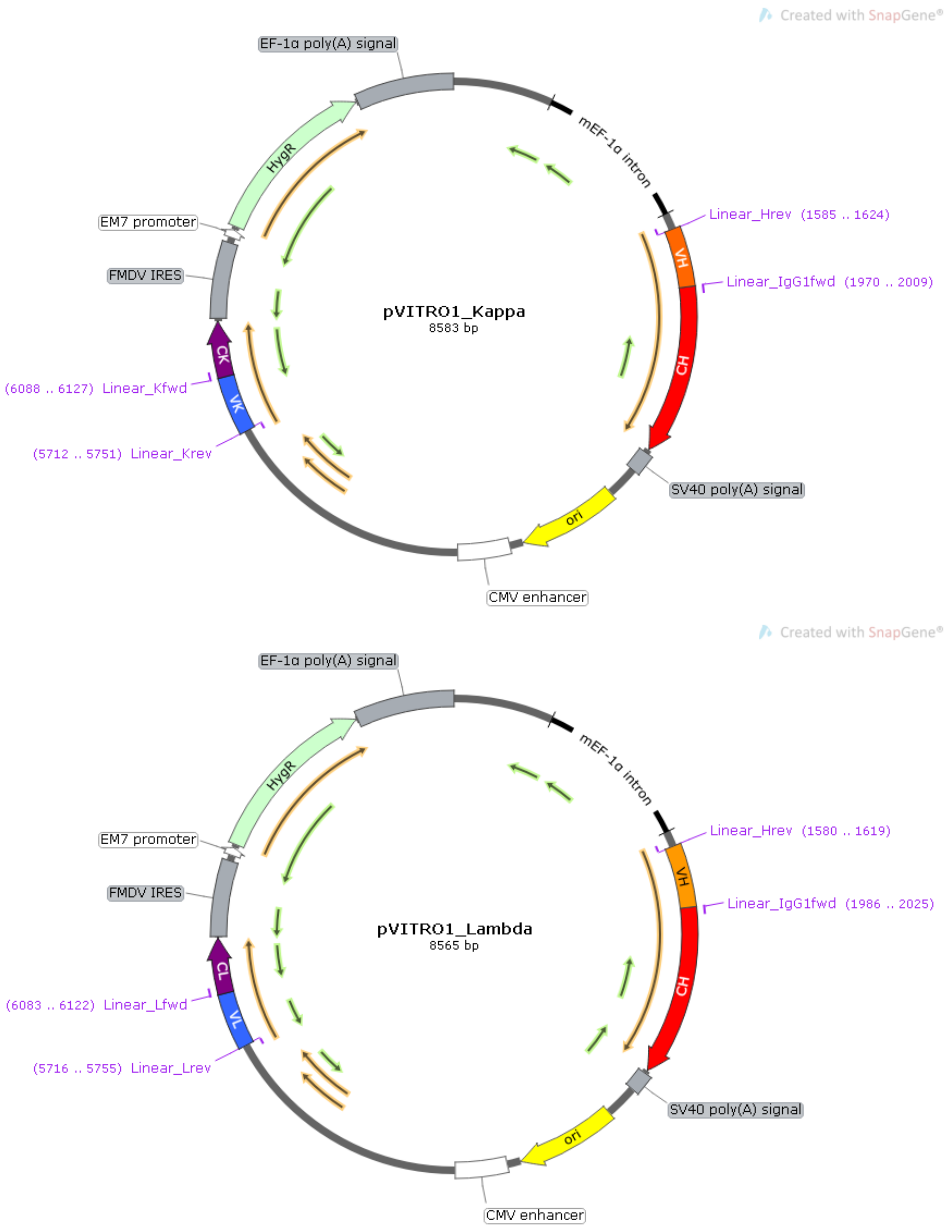


Figure A-1: pVITRO1 IgG kappa and lambda expression plasmids. For pVITRO1_kappa, primer pair Linear_Kfwd and Linear_Hrev amplify one half of the plasmid and primer pair Linear_Krev and Linear_IgG1fwd amplify the other half. For pVITRO1_lambda, primer pair Linear_Lfwd and Linear_Hrev amplify one half of the plasmid and primer pair Linear_Lrev and Linear_IgG1fwd amplify the other half. Heavy and light chain variable regions are inserted upstream of their respective constant regions using PIPE cloning. The plasmid contains a hygromycin resistance gene (*HygR*). Orange arrows indicate ORFs in direct orientation, green arrows indicate ORFs in reverse orientation. pVITRO1 is a high copy number plasmid.

A.2.2 pESL scFv Expression Plasmid

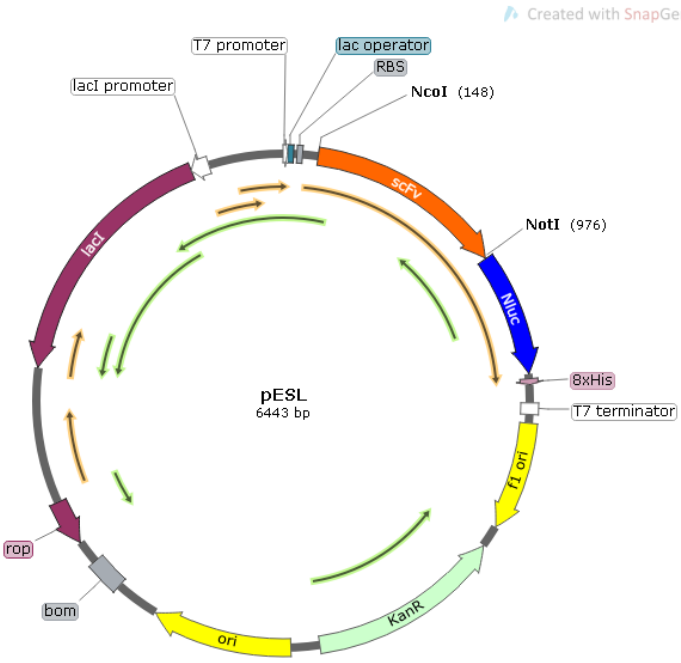


Figure A-2: pESL scFv expression plasmid. scFv are inserted using the NcoI and NotI restriction sites upstream of a NanoLuc luciferase gene and His-tag. Gene expression is under the control of a T7 promoter. The plasmid contains a kanamycin resistance gene (*KanR*) Orange arrows indicate ORFs in direct orientation, green arrows indicate ORFs in reverse orientation. pESL has a pET41 backbone and is a low copy number plasmid.

A.2.3 pHL-sec GP Ectodomain Expression Plasmid

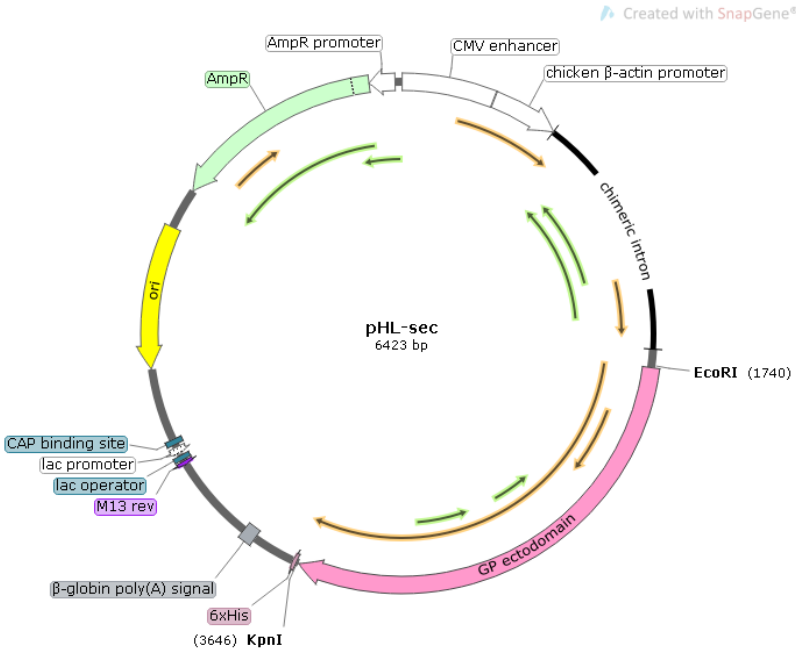


Figure A-3: pHL-sec glycoprotein (GP) expression vector.Filoviridae glycoprotein (GP) ectodomain sequences were inserted using restriction enzymes EcoRI and KpnI. The resulting protein is His-tagged. The plasmid contains an ampicillin resistance gene (*AmpR*). Orange arrows indicate ORFs in direct orientation, green arrows indicate ORFs in reverse orientation. pHL-sec is a high copy number plasmid.

A.3 Buffer Recipes

Binding Buffer (GP purification)

Sodium Phosphate Buffer	50 ml
2 M NaCl	62.5 ml
dH_2O	to 250 ml
pH	7.4

Filter (0.22 µm)

Coomassie Blue (western blot)

Methanol	125 ml
Acetic acid	25 ml
Coomassie Blue G-250	0.5 g
dH_2O	100 ml

Coomassie Destain (western blot)

Methanol	50 ml
Glacial acetic acid	75 ml
dH_2O	to 1 L

Note: add 800 ml dH_2O to bottle before adding the methanol and acetic acid

Elution Buffer (300 mM) (GP purification)

Binding Buffer (filtered)	17 ml
2 M Imidazole (filtered)	3 ml
Final imidazole concentration:	300 mM

Prepare immediately before use

Elution Buffer (500 mM) (GP purification)

Binding Buffer (filtered)	15 ml
2 M Imidazole (filtered)	5 ml
Final Imidazole concentration:	500 mM

Prepare immediately before use

2 M Imidazole (GP purification)

Imidazole ($C_3H_4N_2$)	68.077 g
dH2O	to 500 ml
pH	7.4

Filter (0.22 µm)

6X Loading Buffer (SDS-PAGE)

For non-reducing gel:

4X Stacking*	7 ml
Glycerol	3 ml
SDS ($NaC_{12}H_{25}SO_4$)	1 g
Bromophenol blue	3.2 mg

For reducing gel:
Add 600 µl β-mercaptoethanol to the above or add 1 µl dithiothreitol (DTT) to each protein sample as needed.

*4X Stacking:

Tris Base ($C_4H_{11}NO_3$)	6.05 g
dH_2O	to 40 ml
pH	6.8

Store at -20°C

2 M NaCl (GP purification)

NaCl	58.44 g
dH_2O	to 500 ml

Filter (0.22 µm)

1 M Na_2HPO_4 (GP purification)

Na_2HPO_4	8.9 g
dH_2O	to 250 ml

Filter (0.22 µm)

1 M NaH_2PO_4 (GP purification)

Na_2HPO_4	6.9 g
dH_2O	to 250 ml

Filter (0.22 µm)

Nickel Buffer (GP purification)

$NiSO_4.6H_2O$	6.57 g
dH_2O	250 ml
Final $NiSO_4$ concentration:	0.1 M

Filter (0.22 µm)

SDS-PAGE gel (western blot)

This recipe makes two 0.75 mm gels

Separating gel	8%	10%	12%
dH_2O	5.5	5	4.5 ml
1.5 M Tris (pH 8.8)	2.5	2.5	2.5 ml
Acrylamide (40%)	2	2.5	3 ml
TEMED	10	10	10 µl
SDS (10%)	100	100	100 µl
APS (10%)	50	50	50 µl

Layer absolute ethanol on top of gel until set for straight edge

Stacking gel

dH_2O	3.25 ml
0.5 M Tris (pH 6.8)	1.25 ml
Acrylamide (40%)	465 µl
TEMED	5 µl
SDS (10%)	50 µl
APS (10%)	25 µl

10X SDS Running Buffer (western blot)

Tris Base ($C_4H_{11}NO_3$)	30.35 g
Glycine ($C_2H_5NO_2$)	288.26 g
SDS ($NaC_{12}H_{25}SO_4$)	10 g
dH_2O	to 1 L
pH	8.3

May need gentle heating to dissolve

For 1X solution dilute 1/10 with dH_2O

Sodium Phosphate Buffer (GP purification)

1 M Na_2HPO_4	38.7 ml
1 M NaH_2PO_4	11.3 ml
dH_2O	to 500 ml
Final sodium phosphate concentration:	0.1 M
pH	7.4

Filter (0.22 µm)

50X TAE Buffer

Tris Base ($C_4H_{11}NO_3$)	242 g
EDTA ($C_{10}H_{14}N_2Na_2O_8.2H_2O$)	14.6 g
Glacial acetic acid (CH_3COOH)	57 ml
dH_2O	to 1 L
pH	8.2-8.3

For 1X solution dilute 1/50 with dH_2O

10X Transfer Buffer (western blot)

Tris Base ($C_4H_{11}NO_3$)	30.35 g
Glycine ($C_2H_5NO_2$)	144.13 g
dH_2O	to 1 L
pH	8.3

May need gentle heating to dissolve

For 1X solution dilute 1/10 with dH_2O

Washing Buffer (GP purification)

Binding Buffer (filtered)	49 ml
2 M Imidazole (filtered)	1 ml
Final imidazole concentration:	40 mM

Prepare immediately before use

A.4 Number of Sequences in B cell Development Dataset

Table A-19: Number of heavy chain sequences in the full B cell Development Dataset. Only a single, modal, sequence of each clone is counted.

Donor ID	Gender	Age	Pre-B	Immature	Transitional	Naïve	Total
122	F	24	1290	285	343	259	2177
159	F	28	830	1880	502	349	3561
138	M	41	2629	76	195	497	3397
118	M	48	2149	478	294	647	3568
146	F	50	3832	675	464	245	5216
103	M	52	87	364	65	1606	2122
141	F	65	722	1485	979	501	3687
119	F	68	2966	1704	908	367	5945
160	M	70	89	191	98	3109	3487
111	F	71	620	74	19	1107	1820
149	M	72	1120	1135	178	146	2579
140	F	86	731	436	340	571	2078
		Total	17065	8783	4385	9404	39637

Table A-20: Number of heavy chain sequences in the clean B cell Development Dataset. Only a single, modal, sequence of each clone is counted.

Donor ID	Gender	Age	Pre-B	Immature	Transitional	Naïve	Total
122	F	24	934	220	274	202	1630
159	F	28	632	1339	379	260	2610
138	M	41	2042	54	154	378	2628
118	M	48	1310	391	189	423	2313
146	F	50	2876	506	351	172	3905
103	M	52	69	280	56	1241	1646
141	F	65	553	1081	577	359	2570
119	F	68	1944	1305	678	245	4172
160	M	70	62	149	81	2351	2643
111	F	71	498	59	13	810	1380
149	M	72	888	813	125	105	1931
140	F	86	577	309	269	433	1588
		Total	12385	6506	3146	6979	29016

Table A-21: Number of light chain sequences in the full B cell Development Dataset. Only a single, modal, sequence of each clone is counted.

Donor			Immature		Transitional		Naïve		
ID	Gender	Age	Kappa	Lambda	Kappa	Lambda	Kappa	Lambda	Total
122	F	24	1380	308	349	52	470	87	2646
159	F	28	1292	349	419	140	532	135	2867
138	M	41	255	22	130	125	961	0	1493
107	M	43	0	0	1476	686	163	152	2477
120	M	44	0	0	841	568	378	212	1999
118	M	49	464	348	249	88	281	58	1488
146	F	50	948	516	485	109	131	37	2226
103	F	52	371	164	15	3	554	307	1414
128	F	52	567	400	360	234	0	0	1561
126	M	53	259	83	664	242	1651	897	3796
141	F	65	2021	1414	475	233	478	73	4694
105	F	67	288	170	183	83	417	303	1444
119	F	68	2175	1609	666	27	334	162	4973
160	M	70	304	207	130	71	757	37	1506
111	F	71	188	79	151	67	708	644	1837
149	M	72	848	689	0	0	246	171	1954
132	M	76	0	0	896	381	515	267	2059
162	M	78	0	0	1139	1	1	249	1390
140	F	86	553	368	279	174	764	382	2520
		Total	11913	6726	8907	3284	9341	4173	44344

Table A-22: Number of light chain sequences in the clean B cell Development Dataset. Only a single, modal, sequence of each clone is counted.

Donor			Immature		Transitional		Naïve		
ID	Gender	Age	Kappa	Lambda	Kappa	Lambda	Kappa	Lambda	Total
122	F	24	958	213	194	31	312	47	1755
159	F	28	819	198	266	82	397	103	1865
138	M	41	130	18	101	103	678	0	1030
107	M	43	0	0	1063	438	144	124	1769
120	M	44	0	0	537	360	304	166	1367
118	M	49	315	222	176	64	204	37	1018
146	F	50	620	292	325	52	94	20	1403
103	F	52	256	101	8	2	424	213	1004
128	F	52	402	263	204	120	0	0	989
126	M	53	123	49	332	121	1102	565	2292
141	F	65	1342	869	358	137	339	42	3087
105	F	67	186	89	142	59	333	247	1056
119	F	68	1399	1002	512	15	267	101	3296
160	M	70	175	101	102	52	622	28	1080
111	F	71	99	51	116	38	558	414	1276
149	M	72	552	374	0	0	177	115	1218
132	M	76	0	0	519	219	404	180	1322
162	M	78	0	0	770	1	1	189	961
140	F	86	371	206	192	96	547	247	1659
		Total	7747	4048	5917	1990	6907	2838	29447

A.5 Number of Sequences in Ebola Caucasian Aid Worker Dataset

Table A-23: Number of heavy chain sequences in full Caucasian Aid Worker Dataset. Only a single, modal, sequence of each clone is counted.

Group	Donor ID	Gender	Age	IgA	IgG	IgM	Total
Ebola Survivors	215	M	29	2910	3821	1745	8476
	216	M	34	2009	2257	1427	5693
	217	F	39	554	1177	1132	2863
YF Vacc. Controls Day 0	193	M	27	369	388	435	1192
	199	M	34	597	322	324	1243
	207	F	36	231	289	519	1039
YF Vacc. Controls Day 28	193	M	27	291	778	1106	2175
	199	M	34	310	273	118	701
	207	F	36	308	126	599	1033
			Total	7579	9431	7405	24415

Table A-24: Number of heavy chain sequences in clean Caucasian Aid Worker Dataset. Only a single, modal, sequence of each clone is counted.

Group	Donor ID	Gender	Age	IgA	IgG	IgM	Total
Ebola Survivors	215	M	29	2309	3013	1368	6690
	216	M	34	1577	1796	1131	4504
	217	F	39	429	903	900	2232
YF Vacc. Controls Day 0	193	M	27	285	305	327	917
	199	M	34	480	251	254	985
	207	F	36	182	222	408	812
YF Vacc. Controls Day 28	193	M	27	225	588	867	1680
	199	M	34	250	219	98	567
	207	F	36	216	94	471	781
			Total	5953	7391	5824	19168

Table A-25: Number of light chain sequences in full Caucasian Aid Worker Dataset. Only a single, modal, sequence of each clone is counted.

Group	Donor ID	Gender	Age	Kappa	Lambda	Total
Ebola Survivors	215	M	29	3992	1420	5412
	216	M	34	3606	1195	4801
	217	F	39	3663	406	4069
YF Vacc. Controls Day 0	193	M	27	2133	347	2480
	199	M	34	1964	339	2303
	207	F	36	1800	291	2091
YF Vacc. Controls Day 28	193	M	27	1543	756	2299
	199	M	34	1223	325	1548
	207	F	36	443	472	915
			Total	20367	5551	25918

Table A-26: Number of light chain sequences in clean Caucasian Aid Worker Dataset. Only a single, modal, sequence of each clone is counted.

Group	Donor ID	Gender	Age	Kappa	Lambda	Total
Ebola Survivors	215	M	29	2843	1182	4025
	216	M	34	2878	975	3853
	217	F	39	2893	337	3230
YF Vacc. Controls Day 0	193	M	27	1607	293	1900
	199	M	34	1552	274	1826
	207	F	36	1396	222	1618
YF Vacc. Controls Day 28	193	M	27	1154	585	1739
	199	M	34	920	263	1183
	207	F	36	339	374	713
			Total	15582	4505	20087

A.6 Number of Sequences in Ebola African Survivor Dataset

Table A-27: Number of heavy chain sequences in full African Survivor Dataset. Only a single, modal, sequence of each clone is counted.

Group	Donor ID	IgA	IgG	IgM	Total
Time Point 1	2	1013	1393	958	3364
	50	520	729	261	1510
	14	654	635	215	1504
	82	436	698	416	1550
	15	685	855	412	1952
Time Point 2	2	826	924	131	1881
	50	842	645	117	1604
	14	730	669	187	1586
	82	338	535	413	1286
	15	758	774	465	1997
	Total	6802	7857	3575	18234

Table A-28: Number of heavy chain sequences in clean African Survivor Dataset. Only a single, modal, sequence of each clone is counted.

Group	Donor ID	IgA	IgG	IgM	Total
Time Point 1	2	969	1335	913	3217
	50	501	702	249	1452
	14	631	616	205	1452
	82	419	671	393	1483
	15	657	822	394	1873
Time Point 2	2	799	891	125	1815
	50	807	609	112	1528
	14	709	640	174	1523
	82	331	514	390	1235
	15	735	744	448	1927
	Total	6558	7544	3403	17505

A.7 Number of Sequences in Ribosome Display Library Dataset

Table A-29: Number of scFv sequences in full Ribosome Display Library dataset. Only a single, modal, sequence of each clone is counted.

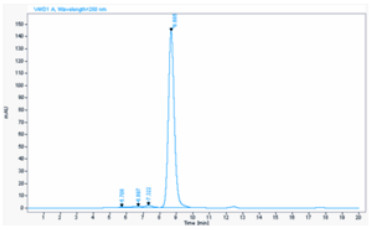
Group	Donor ID	IgH	IgK	IgL	Total
PCR3 Input	215	12716	2757	3407	18880
	217	7177	2989	2944	13110
PCR4 Output	215	474	275	72	821
	217	218	229	47	494
	Total	20585	6250	6470	33305

Table A-30: Number of scFv sequences in clean Ribosome Display Library dataset. Only a single, modal, sequence of each clone is counted.

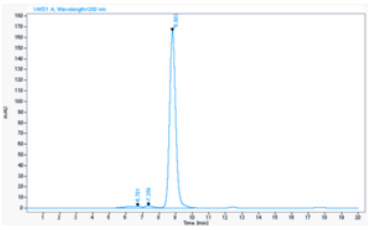
Group	Donor ID	IgH	IgK	IgL	Total
PCR3 Input	215	11689	2224	3078	16991
	217	6620	2570	2645	11835
PCR4 Output	215	428	209	53	690
	217	192	173	40	405
	Total	18929	5176	5816	29921

A.8 HP-SEC Traces of Purified IgG1 Antibodies

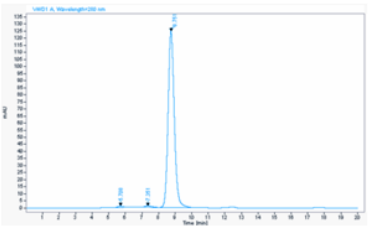
CT0001 (96.50%)



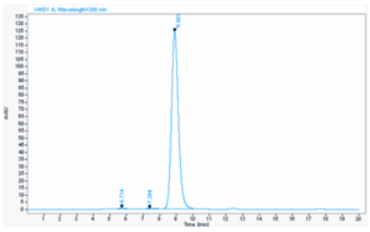
CT0002 (96.01%)



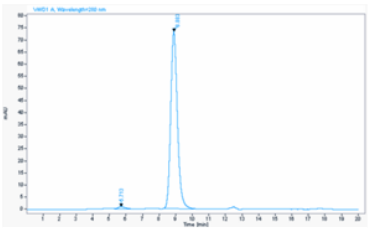
CT0003 (98.50%)



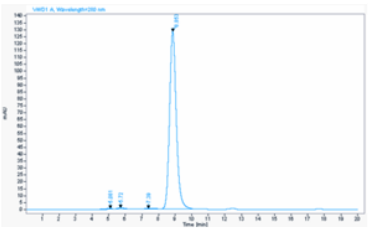
CT0004 (98.83%)



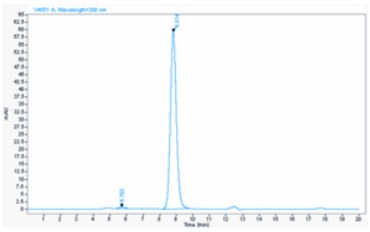
CT0005 (98.63%)



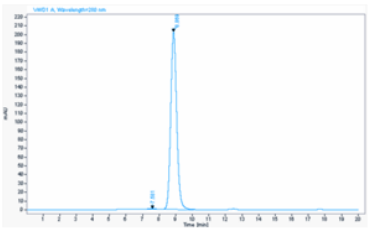
CT0006 (97.89%)



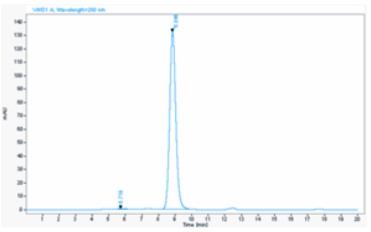
CT0007 (99.04%)



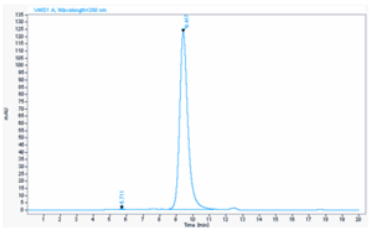
CT0008 (99.33%)



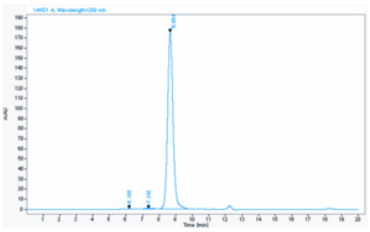
CT0009 (99.39%)



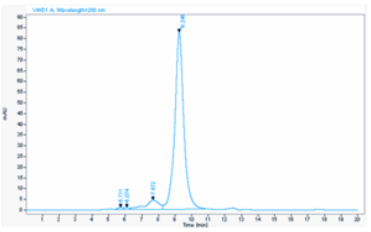
CT0010 (99.68%)



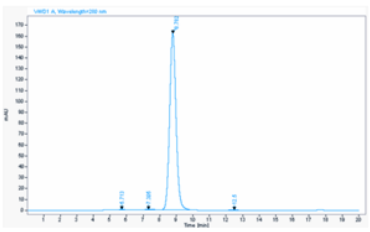
CT0011 (99.17%)



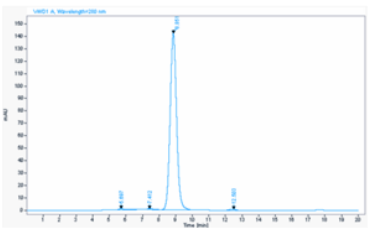
CT0012 (91.71%)



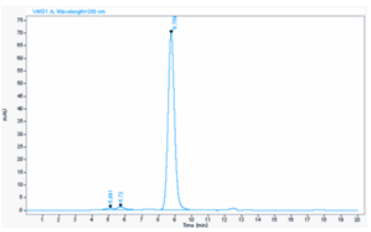
CT0013 (98.70%)



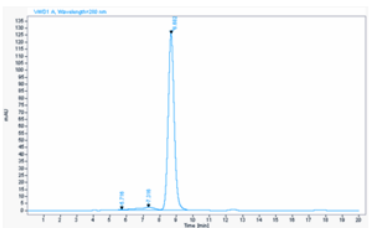
CT0014 (98.07%)



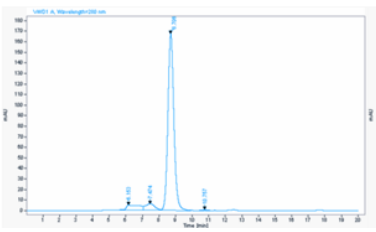
CT0015 (97.06%)



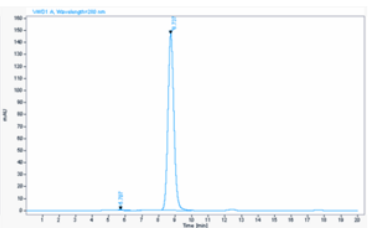
CT0016 (95.13%)



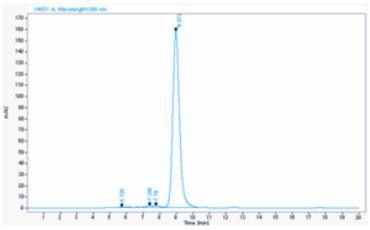
CT0017 (88.12%)



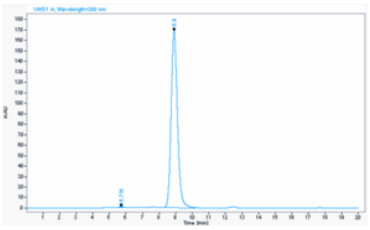
CT0018 (99.61%)



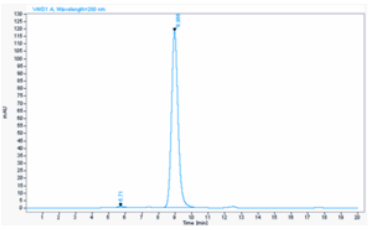
CT0019 (96.80%)



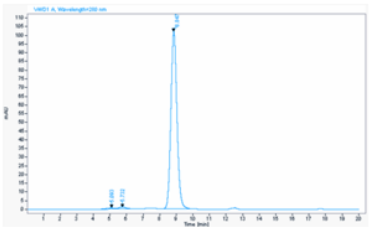
CT0020 (99.58%)



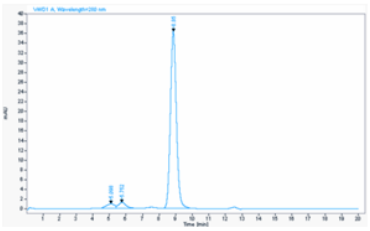
CT0021 (99.18%)



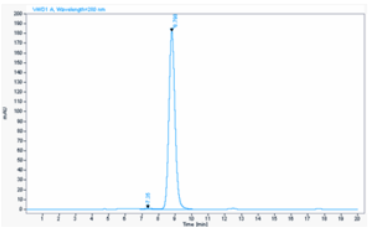
CT0022 (97.79%)



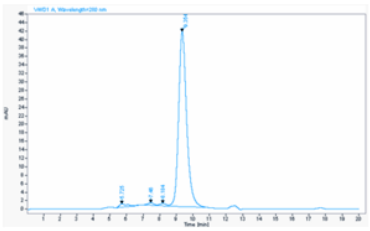
CT0023 (92.93%)



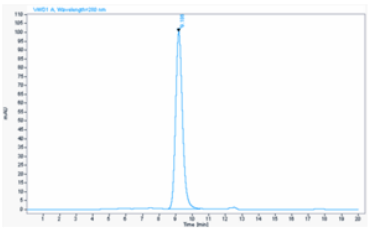
CT0024 (99.22%)



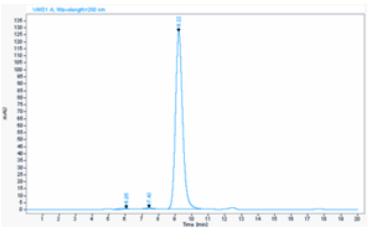
CT0025 (96.41%)



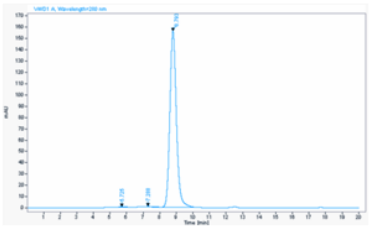
CT0026 (100.00%)



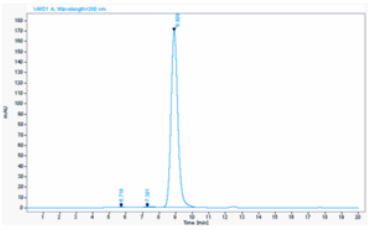
CT0027 (98.03%)



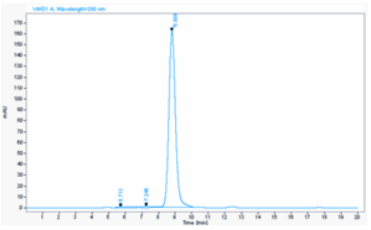
CT0028 (99.02%)



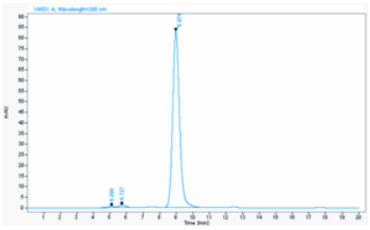
CT0029 (99.19%)



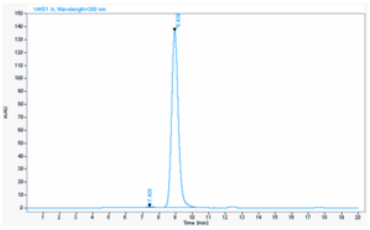
CT0030 (97.46%)



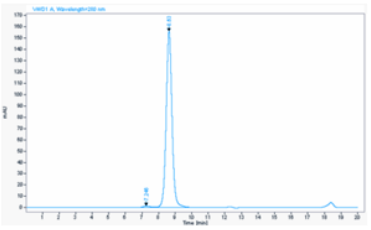
CT0031 (97.63%)



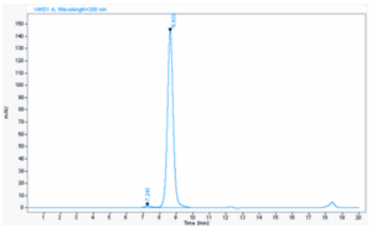
CT0032 (99.44%)



CT0033 (99.10%)



CT0034 (99.06%)



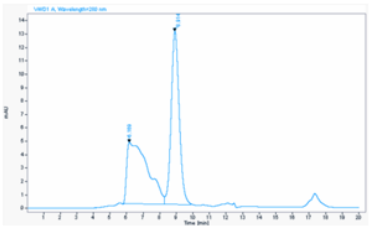
CT0035

IgG could not be expressed

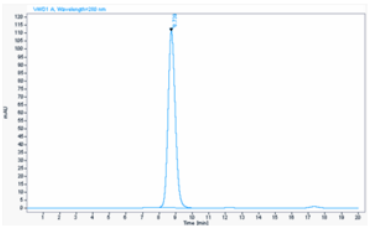
CT0036

IgG could not be expressed

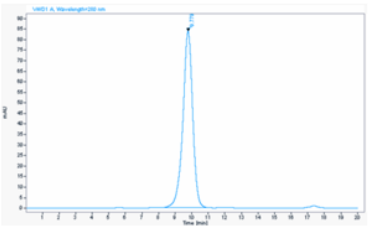
CT0037 (51.72%)



CT0038 (100.00%)



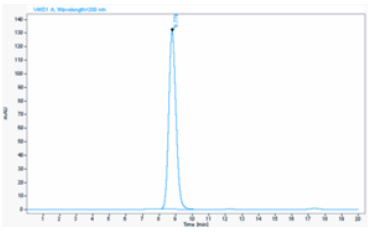
CT0039 (100.00%)



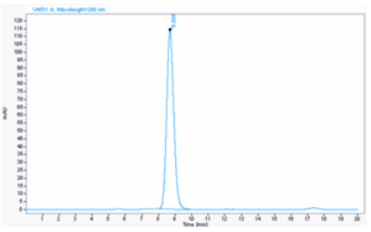
CT0040

IgG could not be expressed

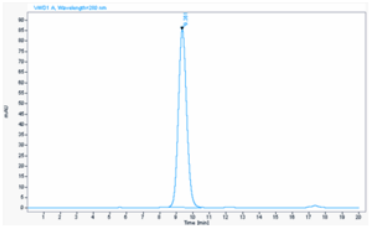
CT0041 (100.00%)



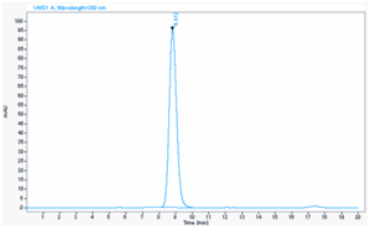
CT0042 (100.00%)



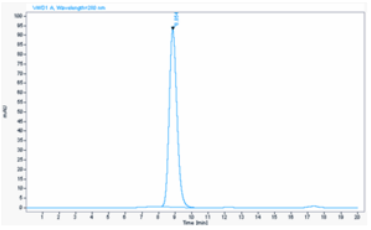
CT0043 (100.00%)



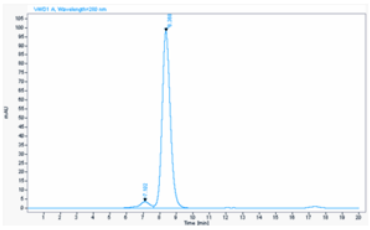
CT0044 (100.00%)



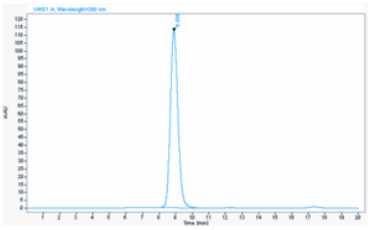
CT0045 (100.00%)



CT0046 (95.52%)



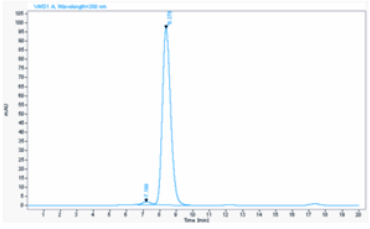
CT0047 (100.00%)



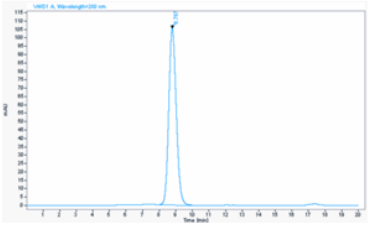
CT0048

IgG could not be expressed

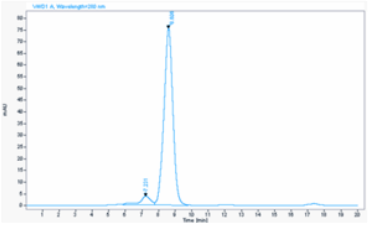
CT0049 (98.43%)



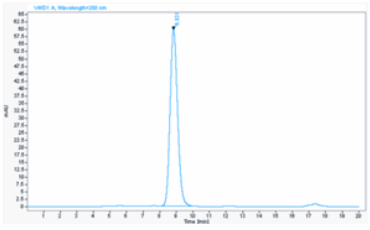
CT0050 (100.00%)



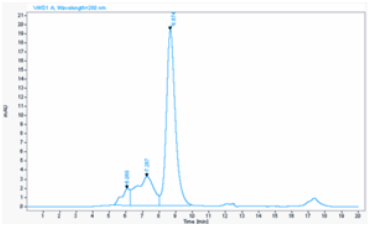
CT0051 (93.94%)



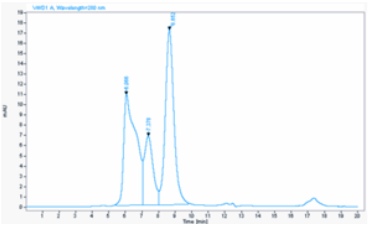
CT0052 (100.00%)



CT0053 (71.34%)



CT0054 (43.23%)



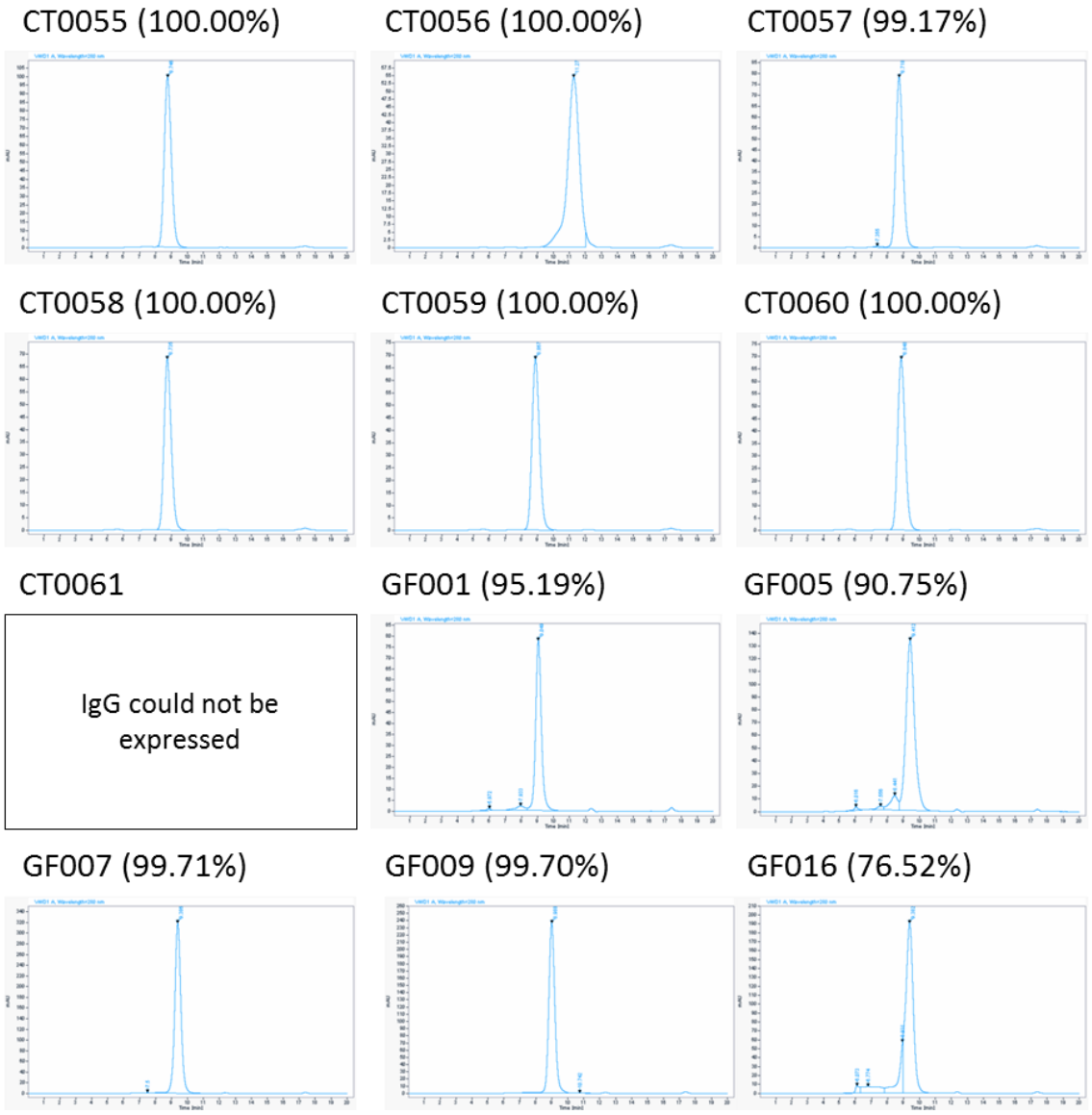


Figure A-4: HP-SEC traces of purified antibodies. Percentages next to antibody name indicate the proportion of the purified antibody that was in monomeric form. Peaks to the left of the large monomer peak indicate aggregated antibody. Peaks to the right of the large monomer peak indicate fragmented antibody.[105]

[105] Antibody purification and HP-SEC was performed by the IgG team at MedImmune (Cambridge, United Kingdom)

A.9 Optimal ELISA Antigen Coating Concentration Titration Curves

A.9.1 Solid-Phase ELISA

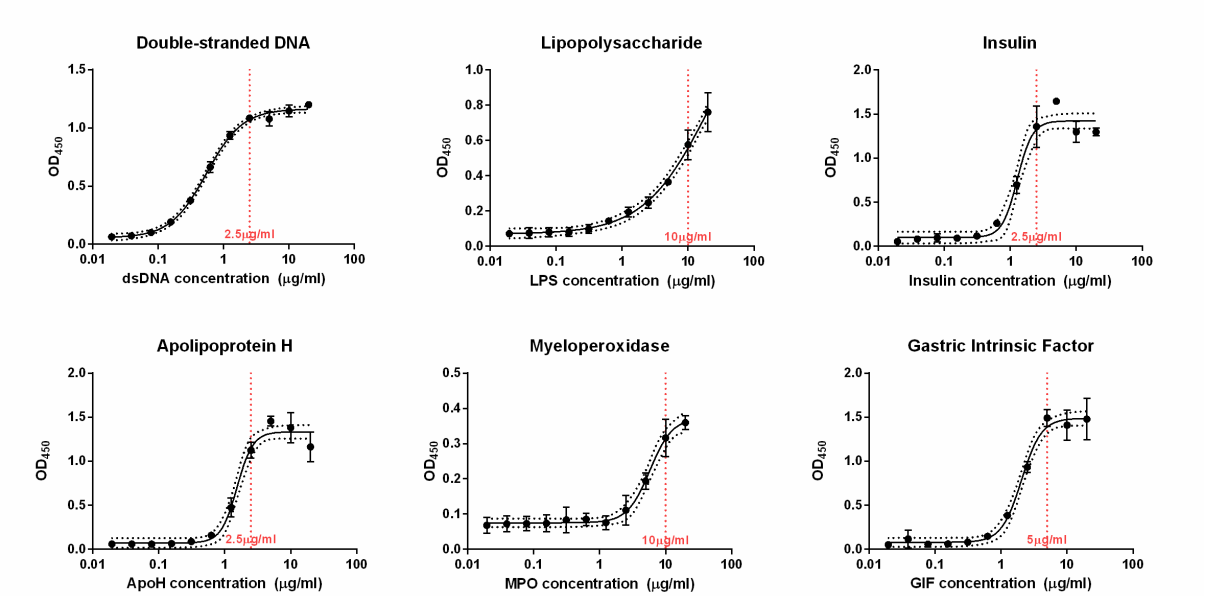


Figure A-5: Solid-phase ELISA antigen coating concentrations. To confirm that all control antibodies bound their corresponding antigen and to determine the optimal antigen coating concentrations, titration curves were conducted for each antigen (Table 2-34, page 111) using its corresponding control antibody (Table 2-35, page 111). The ELISA was blocked with 3% (w/v) skimmed milk powder in PBS and both primary and secondary antibodies were diluted in 3% (w/v) skimmed milk powder in PBS. Antibodies were incubated for 1 h at room temperature. The optimal coating concentration was the concentration at which the curve plateaued or 10µg/ml, whichever was lower.Soluble ELISA using Biotinylated Insulin

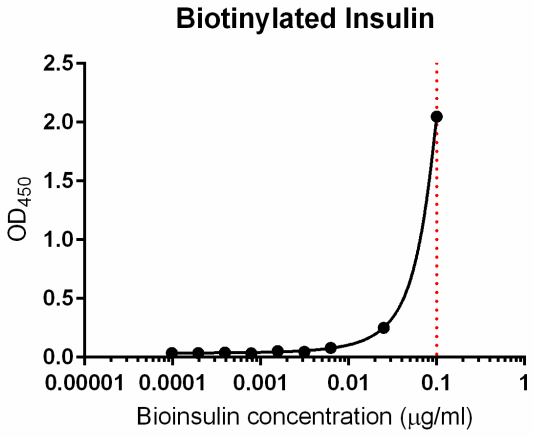


Figure A-6: Soluble ELISA biotinylated insulin coating concentration. To confirm that the anti-insulin control antibody (Table 2-34, page 111) bound biotinylated insulin in the soluble ELISA format, and to determine the optimal biotinylated insulin coating concentration, a titration curve was performed. ELISA was blocked with 3% (w/v) skimmed milk powder in PBS and was incubated at room temperature for 1 h at 800rpm. Readings at 0.05µg/ml and 0.0125µg/ml were excluded from the standard curve as they were outliers (magnetic beads probably lost during washes).

References

References

Aalberse, R. C., Stapel, S. O., Schuurman, J. & Rispens, T. 2009. Immunoglobulin G4: an odd antibody. *Clinical and Experimental Allergy,* 39**,** 469-77.

Abboud, N., Chow, S. K., Saylor, C., Janda, A., Ravetch, J. V., Scharff, M. D. & Casadevall, A. 2010. A requirement for FcγR in antibody-mediated bacterial toxin neutralization. *Journal of Experimental Medicine,* 207**,** 2395-405.

Abhinandan, K. R. & Martin, A. C. 2008. Analysis and improvements to Kabat and structurally correct numbering of antibody variable domains. *Molecular Immunology,* 45**,** 3832-9.

Agrawal, S., Smith, S. A., Tangye, S. G. & Sewell, W. A. 2013. Transitional B cell subsets in human bone marrow. *Clinical and Experimental Immunology,* 174**,** 53-9.

Akamatsu, Y., Tsurushita, N., Nagawa, F., Matsuoka, M., Okazaki, K., Imai, M. & Sakano, H. 1994. Essential residues in V(D)J recombination signals. *Journal of Immunology,* 153**,** 4520-9.

Al-Lazikani, B., Lesk, A. M. & Chothia, C. 1997. Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology,* 273**,** 927-948.

Alamyar, E., Duroux, P., Lefranc, M. P. & Giudicelli, V. 2012a. IMGT(®) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods in Molecular Biology,* 882**,** 569-604.

Alamyar, E., Giudicelli, V., Li, S., Duroux, P. & Lefranc, M. P. 2012b. IMGT/HIGHV-QUEST: The IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Research,* 8**,** 2.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research,* 25**,** 3389-402.

Anand, N. N., Mandal, S., MacKenzie, C. R., Sadowska, J., Sigurskjold, B., Young, N. M., Bundle, D. R. & Narang, S. A. 1991. Bacterial expression and secretion of various single-chain Fv genes encoding proteins specific for a Salmonella serotype B O-antigen. *Journal of Biological Chemistry,* 266**,** 21874-9.

Andrews, S. F., Huang, Y., Kaur, K., Popova, L. I., Ho, I. Y., Pauli, N. T., Henry Dunand, C. J., Taylor, W. M., Lim, S., Huang, M., Qu, X., Lee, J. H., Salgado-Ferrer, M., Krammer, F., Palese, P., Wrammert, J., Ahmed, R. & Wilson, P. C. 2015. Immune history profoundly affects broadly protective B cell responses to influenza. *Science Translational Medicine,* 7**,** 316ra192.

Arbuckle, M. R., McClain, M. T., Rubertone, M. V., Scofield, R. H., Dennis, G. J., James, J. A. & Harley, J. B. 2003. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *New England Journal of Medicine,* 349**,** 1526-33.

Aricescu, A. R., Lu, W. & Jones, E. Y. 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D-Biological Crystallography,* 62**,** 1243-50.

Arpin, C., de Bouteiller, O., Razanajaona, D., Fugier-Vivier, I., Brière, F., Banchereau, J., Lebecque, S. & Liu, Y.-J. 1998. The Normal Counterpart of IgD Myeloma Cells in Germinal

Center Displays Extensively Mutated IgVH Gene, Cμ–Cδ Switch, and λ Light Chain Expression. *The Journal of Experimental Medicine,* 187**,** 1169-1178.

Avnir, Y., Watson, C. T., Glanville, J., Peterson, E. C., Tallarico, A. S., Bennett, A. S., Qin, K., Fu, Y., Huang, C.-Y., Beigel, J. H., Breden, F., Zhu, Q. & Marasco, W. A. 2016. IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. *Scientific Reports,* 6**,** 20842.

Baize, S., Leroy, E. M., Georges-Courbot, M. C., Capron, M., Lansoud-Soukate, J., Debré, P., Fisher-Hoch, S. P., McCormick, J. B. & Georges, A. J. 1999. Defective humoral responses and extensive intravascular apoptosis are associated with fatal outcome in Ebola virus-infected patients. *Nature Medicine,* 5**,** 423-6.

Bakema, J. E. & van Egmond, M. 2011. Immunoglobulin A: A next generation of therapeutic antibodies? *MAbs,* 3**,** 352-61.

Baranzini, S. E., Jeong, M. C., Butunoi, C., Murray, R. S., Bernard, C. C. & Oksenberg, J. R. 1999. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. *Journal of Immunology,* 163**,** 5133-44.

Barrett, D. J. & Ayoub, E. M. 1986. IgG2 subclass restriction of antibody to pneumococcal polysaccharides. *Clinical and Experimental Immunology,* 63**,** 127-34.

Baseler, L., Chertow, D. S., Johnson, K. M., Feldmann, H. & Morens, D. M. 2017. The Pathogenesis of Ebola Virus Disease. *Annual Review of Pathology,* 12**,** 387-418.

Bashford-Rogers, R. J. M., Smith, K. G. C. & Thomas, D. C. 2018. Antibody repertoire analysis in polygenic autoimmune diseases. *Immunology*.

Benckert, J., Schmolka, N., Kreschel, C., Zoller, M. J., Sturm, A., Wiedenmann, B. & Wardemann, H. 2011. The majority of intestinal IgA^+ and IgG^+ plasmablasts in the human gut are antigen-specific. *Journal of Clinical Investigation,* 121**,** 1946-1955.

Berkowska, M. A., Driessen, G. J. A., Bikos, V., Grosserichter-Wagener, C., Stamatopoulos, K., Cerutti, A., He, B., Biermann, K., Lange, J. F., van der Burg, M., van Dongen, J. J. M. & van Zelm, M. C. 2011. Human memory B cells originate from three distinct germinal center-dependent and -independent maturation pathways. *Blood,* 118**,** 2150-2158.

Berkowska, M. A., Schickel, J. N., Grosserichter-Wagener, C., de Ridder, D., Ng, Y. S., van Dongen, J. J. M., Meffre, E. & van Zelm, M. C. 2015. Circulating Human $CD27^-IgA^+$ Memory B Cells Recognize Bacteria with Polyreactive Igs. *Journal of Immunology,* 195**,** 1417-1426.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. 2000. The Protein Data Bank. *Nucleic Acids Research,* 28**,** 235-242.

Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M., Lee, T., Pope, S. H., Riordan, G. S. & Whitlow, M. 1988. Single-chain antigen-binding proteins. *Science,* 242**,** 423-6.

Bischof, J. & Ibrahim, S. M. 2016. bcRep: R Package for Comprehensive Analysis of B Cell Receptor Repertoire Data. *PLoS One,* 11**,** e0161569.

Boc, A., Diallo, A. B. & Makarenkov, V. 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Research,* 40**,** W573-W579.

Boctor, F. N. & Peter, J. B. 1990. IgG subclasses in human chronic schistosomiasis: over-production of schistosome-specific and non-specific IgG4. *Clinical and Experimental Immunology,* 82**,** 574-8.

Boman, H. G. 2003. Antibacterial peptides: basic facts and emerging concepts. *Journal of Internal Medicine,* 254**,** 197-215.

Bond, C. J., Wiesmann, C., Marsters, J. C., Jr. & Sidhu, S. S. 2005. A structure-based database of antibody variable domain diversity. *Journal of Molecular Biology,* 348**,** 699-709.

Bornholdt, Z. A., Ndungo, E., Fusco, M. L., Bale, S., Flyak, A. I., Crowe, J. E., Jr., Chandran, K. & Saphire, E. O. 2016a. Host-Primed Ebola Virus GP Exposes a Hydrophobic NPC1 Receptor-Binding Pocket, Revealing a Target for Broadly Neutralizing Antibodies. *MBio,* 7**,** e02154-15.

Bornholdt, Z. A., Turner, H. L., Murin, C. D., Li, W., Sok, D., Souders, C. A., Piper, A. E., Goff, A., Shamblin, J. D., Wollen, S. E., Sprague, T. R., Fusco, M. L., Pommert, K. B., Cavacini, L. A., Smith, H. L., Klempner, M., Reimann, K. A., Krauland, E., Gerngross, T. U., Wittrup, K. D., Saphire, E. O., Burton, D. R., Glass, P. J., Ward, A. B. & Walker, L. M. 2016b. Isolation of potent neutralizing antibodies from a survivor of the 2014 Ebola virus outbreak. *Science,* 351**,** 1078-83.

Bournazos, S., Klein, F., Pietzsch, J., Seaman, Michael S., Nussenzweig, Michel C. & Ravetch, Jeffrey V. 2014. Broadly Neutralizing Anti-HIV-1 Antibodies Require Fc Effector Functions for In Vivo Activity. *Cell,* 158**,** 1243-1253.

Boursier, L., Su, W. & Spencer, J. 2003. Imprint of somatic hypermutation differs in human immunoglobulin heavy and lambda chain variable gene segments. *Molecular Immunology,* 39**,** 1025-34.

Brauninger, A., Goossens, T., Rajewsky, K. & Kuppers, R. 2001. Regulation of immunoglobulin light chain gene rearrangements during early B cell development in the human. *European Journal of Immunology,* 31**,** 3631-3637.

Brezinschek, H. P., Foster, S. J., Dörner, T., Brezinschek, R. I. & Lipsky, P. E. 1998. Pairing of variable heavy and variable kappa chains in individual naive and memory B cells. *Journal of Immunology,* 160**,** 4762-4767.

Briney, B. S., Willis, J. R. & Crowe, J. E. 2012. Human Peripheral Blood Antibodies with Long HCDR3s Are Established Primarily at Original Recombination Using a Limited Subset of Germline Genes. *Plos One,* 7**,** 13.

Buchner, C., Bryant, C., Eslami, A. & Lakos, G. 2014. Anti-Nuclear Antibody Screening Using HEp-2 Cells. *Journal of Visualized Experiments : JoVE***,** 51211.

Bunker, J. J., Erickson, S. A., Flynn, T. M., Henry, C., Koval, J. C., Meisel, M., Jabri, B., Antonopoulos, D. A., Wilson, P. C. & Bendelac, A. 2017. Natural polyreactive IgA antibodies coat the intestinal microbiota. *Science*.

Burk, R., Bollinger, L., Johnson, J. C., Wada, J., Radoshitzky, S. R., Palacios, G., Bavari, S., Jahrling, P. B. & Kuhn, J. H. 2016. Neglected filoviruses. *FEMS Microbiology Reviews,* 40**,** 494-519.

Burkovitz, A. & Ofran, Y. 2016. Understanding differences between synthetic and natural antibodies can help improve antibody engineering. *MAbs,* 8**,** 278-87.

Cantaert, T., Schickel, J. N., Bannock, J. M., Ng, Y. S., Massad, C., Oe, T., Wu, R., Lavoie, A., Walter, J. E., Notarangelo, L. D., Al-Herz, W., Kilic, S. S., Ochs, H. D., Nonoyama, S., Durandy, A. & Meffre, E. 2015. Activation-Induced Cytidine Deaminase Expression in Human B Cell Precursors Is Essential for Central B Cell Tolerance. *Immunity,* 43**,** 884-95.

Chailyan, A., Marcatili, P., Cirillo, D. & Tramontano, A. 2011. Structural repertoire of immunoglobulin lambda light chains. *Proteins-Structure Function and Bioinformatics,* 79**,** 1513-1524.

Chan, K. R., Ong, E. Z., Mok, D. Z. L. & Ooi, E. E. 2015. Fc receptors and their influence on efficacy of therapeutic antibodies for treatment of viral diseases. *Expert Review of Anti-Infective Therapy,* 13**,** 1351-1360.

Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M. & Wittrup, K. D. 2006. Isolating and engineering human antibodies using yeast surface display. *Nature Protocols,* 1**,** 755-68.

Chapman, C. J., Spellerberg, M. B., Smith, G. A., Carter, S. J., Hamblin, T. J. & Stevenson, F. K. 1993. Autoanti-red cell antibodies synthesized by patients with infectious mononucleosis utilize the VH4-21 gene segment. *Journal of Immunology,* 151**,** 1051-61.

Chen, K., Xu, W., Wilson, M., He, B., Miller, N. W., Bengten, E., Edholm, E.-S., Santini, P. A., Rath, P., Chiu, A., Cattalini, M., Litzman, J., Bussel, J., Huang, B., Meini, A., Riesbeck, K., Cunningham-Rundles, C., Plebani, A. & Cerutti, A. 2009. Immunoglobulin D enhances immune surveillance by activating antimicrobial, pro-inflammatory and B cell-stimulating programs in basophils. *Nature Immunology,* 10**,** 889-898.

Chen, Z. J., Wheeler, C. J., Shi, W., Wu, A. J., Yarboro, C. H., Gallagher, M. & Notkins, A. L. 1998. Polyreactive antigen-binding B cells are the predominant cell type in the newborn B cell repertoire. *European Journal of Immunology,* 28**,** 989-994.

Chothia, C. & Lesk, A. M. 1987. Canonical structures for the hypervariable regions of immunoglobulins. *Journal of Molecular Biology,* 196**,** 901-917.

Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smithgill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S., Alzari, P. M. & Poljak, R. J. 1989. Conformations of Immunoglobulin Hypervariable Regions. *Nature,* 342**,** 877-883.

Christley, S., Scarborough, W., Salinas, E., Rounds, W. H., Toby, I. T., Fonner, J. M., Levin, M. K., Kim, M., Mock, S. A., Jordan, C., Ostmeyer, J., Buntzman, A., Rubelt, F., Davila, M. L., Monson, N. L., Scheuermann, R. H. & Cowell, L. G. 2018. VDJServer: A Cloud-Based Analysis Portal and Data Commons for Immune Repertoire Sequences and Rearrangements. *Frontiers in Immunology,* 9**,** 976.

Chughtai, A. A., Barnes, M. & Macintyre, C. R. 2016. Persistence of Ebola virus in various body fluids during convalescence: evidence and implications for disease transmission and control. *Epidemiology and Infection,* 144**,** 1652-1660.

Chui, S. H., Lam, C. W. K. & Lai, K. N. 1990. Light-chain ratios of immunoglobulins G, A, and M determined by enzyme immunoassay. *Clinical Chemistry,* 36**,** 501-502.

Coker, H. A., Harries, H. E., Banfield, G. K., Carr, V. A., Durham, S. R., Chevretton, E., Hobby, P., Sutton, B. J. & Gould, H. J. 2005. Biased use of VH5 IgE-positive B cells in the nasal mucosa in allergic rhinitis. *Journal of Allergy and Clinical Immunology,* 116**,** 445-52.

Corcoran, M. M., Phad, G. E., Bernat, N. V., Stahl-Hennig, C., Sumida, N., Persson, M. A. A., Martin, M. & Hedestam, G. B. K. 2016. Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nature Communications,* 7**,** 13642.

Das, S., Hirano, M., Tako, R., McCallister, C. & Nikolaidis, N. 2012. Evolutionary Genomics of Immunoglobulin-Encoding Loci in Vertebrates. *Current Genomics,* 13**,** 95-102.

Das, S., Nikolaidis, N., Klein, J. & Nei, M. 2008. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 16647-16652.

Deen, G. F., Broutet, N., Xu, W., Knust, B., Sesay, F. R., McDonald, S. L. R., Ervin, E., Marrinan, J. E., Gaillard, P., Habib, N., Liu, H., Liu, W., Thorson, A. E., Yamba, F., Massaquoi, T. A., James, F., Ariyarajah, A., Ross, C., Bernstein, K., Coursier, A., Klena, J., Carino, M., Wurie, A. H., Zhang, Y., Dumbuya, M. S., Abad, N., Idriss, B., Wi, T., Bennett, S. D., Davies, T., Ebrahim, F. K., Meites, E., Naidoo, D., Smith, S. J., Ongpin, P., Malik, T., Banerjee, A., Erickson, B. R., Liu, Y., Liu, Y., Xu, K., Brault, A., Durski, K. N., Winter, J., Sealy, T., Nichol, S. T., Lamunu, M., Bangura, J., Landoulsi, S., Jambai, A., Morgan, O., Wu, G., Liang, M., Su, Q., Lan, Y., Hao, Y., Formenty, P., Stroher, U. & Sahr, F. 2017. Ebola RNA Persistence in Semen of Ebola Virus Disease Survivors - Final Report. *New England Journal of Medicine,* 377**,** 1428-1437.

DeKosky, B. J., Kojima, T., Rodin, A., Charab, W., Ippolito, G. C., Ellington, A. D. & Georgiou, G. 2015. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nature Medicine,* 21**,** 86-91.

DeKosky, B. J., Lungu, O. I., Park, D., Johnson, E. L., Charab, W., Chrysostomou, C., Kuroda, D., Ellington, A. D., Ippolito, G. C., Gray, J. J. & Georgiou, G. 2016. Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** E2636-45.

Díaz-Zaragoza, M., Hernández-Ávila, R., Viedma-Rodríguez, R., Arenas-Aranda, D. & Ostoa-Saloma, P. 2015. Natural and adaptive IgM antibodies in the recognition of tumor-associated antigens of breast cancer (Review). *Oncology Reports,* 34**,** 1106-1114.

Dimitrov, J. D., Planchais, C., Roumenina, L. T., Vassilev, T. L., Kaveri, S. V. & Lacroix-Desmazes, S. 2013. Antibody polyreactivity in health and disease: statu variabilis. *Journal of Immunology,* 191**,** 993-9.

Dodev, T. S., Karagiannis, P., Gilbert, A. E., Josephs, D. H., Bowen, H., James, L. K., Bax, H. J., Beavil, R., Pang, M. O., Gould, H. J., Karagiannis, S. N. & Beavil, A. J. 2014. A tool kit for rapid cloning and expression of recombinant antibodies. *Scientific Reports,* 4**,** 10.

Doorenspleet, M. E., Klarenbeek, P. L., de Hair, M. J., van Schaik, B. D., Esveldt, R. E., van Kampen, A. H., Gerlag, D. M., Musters, A., Baas, F., Tak, P. P. & de Vries, N. 2014. Rheumatoid arthritis synovial tissue harbours dominant B-cell and plasma-cell clones associated with autoreactivity. *Annals of the Rheumatic Diseases,* 73**,** 756-62.

Dörner, T., Farner, N. L. & Lipsky, P. E. 1999. Ig lambda and heavy chain gene usage in early untreated systemic lupus erythematosus suggests intensive B cell stimulation. *Journal of Immunology,* 163**,** 1027-36.

Dörner, T., Foster, S. J., Farner, N. L. & Lipsky, P. E. 1998. Immunoglobulin kappa chain receptor editing in systemic lupus erythematosus. *Journal of Clinical Investigation,* 102**,** 688-694.

Droupadi, P. R. 1994. Mechanism of allergenic cross-reactions--IV. Evidence for participation of aromatic residues in the ligand-binding site of two multi-specific IgE monoclonal antibodies. *Molecular Immunology,* 31**,** 703-703.

Duehr, J., Wohlbold, T. J., Oestereich, L., Chromikova, V., Amanat, F., Rajendran, M., Gomez-Medina, S., Mena, I., tenOever, B. R., Garcia-Sastre, A., Basler, C. F., Munoz-Fontela, C. & Krammer, F. 2017. Novel Cross-Reactive Monoclonal Antibodies against Ebolavirus Glycoproteins Show Protection in a Murine Challenge Model. *Journal of Virology,* 91.

Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., Shi, J. Y. & Deane, C. M. 2014. SAbDab: the structural antibody database. *Nucleic Acids Research,* 42**,** D1140-D1146.

Dunn-Walters, D., Boursier, L. & Spencer, J. 2000. Effect of somatic hypermutation on potential N-glycosylation sites in human immunoglobulin heavy chain variable regions. *Molecular Immunology,* 37**,** 107-13.

Dunn-Walters, D., Townsend, C., Sinclair, E. & Stewart, A. 2018. Immunoglobulin gene analysis as a tool for investigating human immune responses. *Immunological Reviews,* 284**,** 132-147.

Elhanati, Y., Sethna, Z., Marcou, Q., Callan, C. G., Jr., Mora, T. & Walczak, A. M. 2015. Inferring processes underlying B-cell repertoire diversity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370.

Erikson, J., Martinis, J. & Croce, C. M. 1981. Assignment of the genes for human lambda-immunoglobulin chains to chromosome-22. *Nature,* 294**,** 173-175.

Falasca, L., Agrati, C., Petrosillo, N., Di Caro, A., Capobianchi, M. R., Ippolito, G. & Piacentini, M. 2015. Molecular mechanisms of Ebola virus pathogenesis: focus on cell death. *Cell Death and Differentiation,* 22**,** 1250-1259.

Fausther-Bovendo, H., Qiu, X., McCorrister, S., Westmacott, G., Sandstrom, P., Castilletti, C., Di Caro, A., Ippolito, G. & Kobinger, G. P. 2017. Ebola virus infection induces autoimmunity against dsDNA and HSP60. *Scientific Reports,* 7**,** 42147.

Feeney, A. J., Tang, A. & Ogwaro, K. M. 2000. B-cell repertoire formation: role of the recombination signal sequence in non-random V segment utilization. *Immunological Reviews,* 175**,** 59-69.

Fehr, T., Skrastina, D., Pumpens, P. & Zinkernagel, R. M. 1998. T cell-independent type I antibody response against B cell epitopes expressed repetitively on recombinant virus particles. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 9477-9481.

Feldmann, H., Nichol, S. T., Klenk, H. D., Peters, C. J. & Sanchez, A. 1994. Characterization of filoviruses based on differences in structure and antigenicity of the virion glycoprotein. *Virology,* 199**,** 469-73.

Fields, B. A., Goldbaum, F. A., Ysern, X., Poijak, R. J. & Mariuzza, R. A. 1995. Molecular basis of antigen mimicry by an anti-idiotope. *Nature,* 374**,** 739.

Flyak, A. I., Kuzmina, N., Murin, C. D., Bryan, C., Davidson, E., Gilchuk, P., Gulka, C. P., Ilinykh, P. A., Shen, X., Huang, K., Ramanathan, P., Turner, H., Fusco, M. L., Lampley, R., Kose, N., King, H., Sapparapu, G., Doranz, B. J., Ksiazek, T. G., Wright, D. W., Saphire, E. O., Ward, A. B., Bukreyev, A. & Crowe, J. E., Jr. 2018. Broadly neutralizing antibodies from human survivors

target a conserved site in the Ebola virus glycoprotein HR2-MPER region. *Nature Microbiology,* 3**,** 670-677.

Flyak, A. I., Shen, X., Murin, C. D., Turner, H. L., David, J. A., Fusco, M. L., Lampley, R., Kose, N., Ilinykh, P. A., Kuzmina, N., Branchizio, A., King, H., Brown, L., Bryan, C., Davidson, E., Doranz, B. J., Slaughter, J. C., Sapparapu, G., Klages, C., Ksiazek, T. G., Saphire, E. O., Ward, A. B., Bukreyev, A. & Crowe, J. E., Jr. 2016. Cross-Reactive and Potent Neutralizing Antibody Responses in Human Survivors of Natural Ebolavirus Infection. *Cell,* 164**,** 392-405.

Forconi, F., Sozzi, E., Rossi, D., Sahota, S. S., Amato, T., Raspadori, D., Trentin, L., Leoncini, L., Gaidano, G. & Lauria, F. 2008. Selective influences in the expressed immunoglobulin heavy and light chain gene repertoire in hairy cell leukemia. *Haematologica-the Hematology Journal,* 93**,** 697-705.

Francica, J. R., Varela-Rohena, A., Medvec, A., Plesa, G., Riley, J. L. & Bates, P. 2010. Steric Shielding of Surface Epitopes and Impaired Immune Recognition Induced by the Ebola Virus Glycoprotein. *PLOS Pathogens,* 6**,** e1001098.

Francis, J. N., James, L. K., Paraskevopoulos, G., Wong, C., Calderon, M. A., Durham, S. R. & Till, S. J. 2008. Grass pollen immunotherapy: IL-10 induction and suppression of late responses precedes IgG4 inhibitory antibody activity. *Journal of Allergy and Clinical Immunology,* 121**,** 1120-1125.e2.

Frei, J. C., Nyakatura, E. K., Zak, S. E., Bakken, R. R., Chandran, K., Dye, J. M. & Lai, J. R. 2016. Bispecific Antibody Affords Complete Post-Exposure Protection of Mice from Both Ebola (Zaire) and Sudan Viruses. *Scientific Reports,* 6**,** 19193.

Friedensohn, S., Khan, T. A. & Reddy, S. T. 2017. Advanced Methodologies in High-Throughput Sequencing of Immune Repertoires. *Trends in Biotechnology,* 35**,** 203-214.

Frost, S. D. W., Murrell, B., Hossain, A. S. M. M., Silverman, G. J. & Pond, S. L. K. 2015. Assigning and visualizing germline genes in antibody repertoires. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 370**,** 20140240.

Fujii, K., Tsuji, M., Murota, K., Terato, K., Shimozuru, Y. & Nagai, Y. 1989. An improved enzyme-linked immunosorbent assay of anti-collagen antibodies in human serum. *Journal of Immunological Methods,* 124**,** 63-70.

Fulcher, D. A., Lyons, A. B., Korn, S. L., Cook, M. C., Koleda, C., Parish, C., Fazekas de St Groth, B. & Basten, A. 1996. The fate of self-reactive B cells depends primarily on the degree of antigen receptor engagement and availability of T cell help. *Journal of Experimental Medicine,* 183**,** 2313-28.

Gaëta, B. A., Malming, H. R., Jackson, K. J., Bain, M. E., Wilson, P. & Collins, A. M. 2007. iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences. *Bioinformatics,* 23**,** 1580-7.

Galson, J. D., Trück, J., Fowler, A., Münz, M., Cerundolo, V., Pollard, A. J., Lunter, G. & Kelly, D. F. 2015. In-Depth Assessment of Within-Individual and Inter-Individual Variation in the B Cell Receptor Repertoire. *Frontiers in Immunology,* 6.

Gearhart, P. J., Hurwitz, J. L. & Cebra, J. J. 1980. Successive switching of antibody isotypes expressed within the lines of a B-cell clone. *Proceedings of the National Academy of Sciences of the United States of America,* 77**,** 5424-5428.

Gilchuk, P., Kuzmina, N., Ilinykh, P. A., Huang, K., Gunn, B. M., Bryan, A., Davidson, E., Doranz, B. J., Turner, H. L., Fusco, M. L., Bramble, M. S., Hoff, N. A., Binshtein, E., Kose, N., Flyak, A. I., Flinko, R., Orlandi, C., Carnahan, R., Parrish, E. H., Sevy, A. M., Bombardi, R. G., Singh, P. K., Mukadi, P., Muyembe-Tamfum, J. J., Ohi, M. D., Saphire, E. O., Lewis, G. K., Alter, G., Ward, A. B., Rimoin, A. W., Bukreyev, A. & Crowe, J. E., Jr. 2018. Multifunctional Pan-ebolavirus Antibody Recognizes a Site of Broad Vulnerability on the Ebolavirus Glycoprotein. *Immunity,* 49**,** 363-374.

Giudicelli, V., Duroux, P., Lavoie, A., Aouinti, S., Lefranc, M. P. & Kossida, S. 2015. From IMGT-ONTOLOGY to IMGT/HighVQUEST for NGS Immunoglobulin (IG) and T cell Receptor (TR) Repertoires in Autoimmune and Infectious Diseases. *Autoimmune and Infectious Diseases: Open Access,* 1.

Glanville, J., Zhai, W. W., Berka, J., Telman, D., Huerta, G., Mehta, G. R., Ni, I., Mei, L., Sundar, P. D., Day, G. M. R., Cox, D., Rajpal, A. & Pons, J. 2009. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 20216-20221.

González-González, E., Alvarez, M. M., Márquez-Ipiña, A. R., Trujillo-de Santiago, G., Rodríguez-Martínez, L. M., Annabi, N. & Khademhosseini, A. 2017. Anti-Ebola therapies based on monoclonal antibodies: current state and challenges ahead. *Critical Reviews in Biotechnology,* 37**,** 53-68.

Goodnow, C. C., Crosbie, J., Adelstein, S., Lavoie, T. B., Smith-Gill, S. J., Brink, R. A., Pritchard-Briscoe, H., Wotherspoon, J. S., Loblay, R. H., Raphael, K. & et al. 1988. Altered immunoglobulin expression and functional silencing of self-reactive B lymphocytes in transgenic mice. *Nature,* 334**,** 676-82.

Goossens, T., Klein, U. & Küppers, R. 1998. Frequent occurrence of deletions and duplications during somatic hypermutation: Implications for oncogene translocations and heavy chain disease. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 2463-2468.

Gould, H. J. & Sutton, B. J. 2008. IgE in allergy and asthma today. *Nature Reviews Immunology,* 8**,** 205-17.

Gould, H. J., Sutton, B. J., Beavil, A. J., Beavil, R. L., McCloskey, N., Coker, H. A., Fear, D. & Smurthwaite, L. 2003. The Biology of IgE and the Basis of Allergic Disease. *Annual Review of Immunology,* 21**,** 579-628.

Grard, G., Biek, R., Muyembe Tamfum, J.-J., Fair, J., Wolfe, N., Formenty, P., Paweska, J. & Leroy, E. 2011. Emergence of Divergent Zaire Ebola Virus Strains in Democratic Republic of the Congo in 2007 and 2008. *The Journal of Infectious Diseases,* 204**,** S776-S784.

Greenberg, A. S., Steiner, L., Kasahara, M. & Flajnik, M. F. 1993. Isolation of a shark immunoglobulin light chain cDNA clone encoding a protein resembling mammalian kappa light chains - implications for the evolution of light chains. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 10603-10607.

Grönwall, C. & Silverman, G. J. 2014. Natural IgM: Beneficial autoantibodies for the control of inflammatory and autoimmune disease? *Journal of Clinical Immunology,* 34**,** S12-S21.

Gunti, S. & Notkins, A. L. 2015. Polyreactive Antibodies: Function and Quantification. *Journal of Infectious Diseases,* 212**,** S42-S46.

Gupta, N. T., Vander Heiden, J. A., Uduman, M., Gadala-Maria, D., Yaari, G. & Kleinstein, S. H. 2015. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics,* 31**,** 3356-8.

Guruprasad, K., Reddy, B. V. B. & Pandit, M. W. 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering,* 4**,** 155-161.

Güven, E., Duus, K., Lydolph, M. C., Jørgensen, C. S., Laursen, I. & Houen, G. 2014. Non-specific binding in solid phase immunoassays for autoantibodies correlates with inflammation markers. *Journal of Immunological Methods,* 403**,** 26-36.

Halverson, R., Torres, R. M. & Pelanda, R. 2004. Receptor editing is the main mechanism of B cell tolerance toward membrane antigens. *Nature Immunology,* 5**,** 645-650.

Hamerscasterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., Bendahman, N. & Hamers, R. 1993. Naturally occurring antibodies devoid of light chains. *Nature,* 363**,** 446-448.

Hamming, R. W. 1950. Error Detecting and Error Correcting Codes. *Bell System Technical Journal,* 29**,** 147–160.

Hanson, L. Å. & Korotkova, M. 2002. The role of breastfeeding in prevention of neonatal infection. *Seminars in Neonatology,* 7**,** 275-281.

Hartley, S. B., Crosbie, J., Brink, R., Kantor, A. B., Basten, A. & Goodnow, C. C. 1991. Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens. *Nature,* 353**,** 765-9.

He, J.-S., Subramaniam, S., Narang, V., Srinivasan, K., Saunders, S. P., Carbajo, D., Wen-Shan, T., Hidayah Hamadee, N., Lum, J., Lee, A., Chen, J., Poidinger, M., Zolezzi, F., Lafaille, J. J. & Curotto de Lafaille, M. A. 2017. IgG1 memory B cells keep the memory of IgE responses. *Nature Communications,* 8**,** 641.

He, M., Kang, A. S., Hamon, M., Humphreys, A. S., Gani, M. & Taussig, M. J. 1995. Characterization of a progesterone-binding, three-domain antibody fragment (V_H/K) expressed in *Escherichia coli*. *Immunology,* 84**,** 662-8.

Hehle, V., Fraser, L. D., Tahir, R., Kipling, D., Wu, Y. C., Lutalo, P. M. K., Cason, J., Choong, L., D'Cruz, D. P., Cope, A. P., Dunn-Walters, D. K. & Spencer, J. 2015. Immunoglobulin kappa variable region gene selection during early human B cell development in health and systemic lupus erythematosus. *Molecular Immunology,* 65**,** 215-223.

Henderson, K. A., Streltsov, V. A., Coley, A. M., Dolezal, O., Hudson, P. J., Batchelor, A. H., Gupta, A., Bai, T., Murphy, V. J., Anders, R. F., Foley, M. & Nuttall, S. D. 2007. Structure of an IgNAR-AMA1 complex: targeting a conserved hydrophobic cleft broadens malarial strain recognition. *Structure,* 15**,** 1452-66.

Hershberg, U. & Luning Prak, E. T. 2015. The analysis of clonal expansions in normal and autoimmune B cell repertoires. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 370**,** 20140239.

Hershberg, U. & Shlomchik, M. J. 2006. Differences in potential for amino acid change after mutation reveals distinct strategies for kappa and lambda light-chain variation. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 15963-15968.

Hervé, M., Isnardi, I., Ng, Y. S., Bussel, J. B., Ochs, H. D., Cunningham-Rundles, C. & Meffre, E. 2007. CD40 ligand and MHC class II expression are essential for human peripheral B cell tolerance. *Journal of Experimental Medicine,* 204**,** 1583-93.

Hervé, M., Xu, K., Ng, Y. S., Wardemann, H., Albesiano, E., Messmer, B. T., Chiorazzi, N. & Meffre, E. 2005. Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. *Journal of Clinical Investigation,* 115**,** 1636-43.

Hieter, P. A., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. 1981. Human immunoglobulin-kappa light-chain genes are deleted or rearranged in lambda-producing B-cells. *Nature,* 290**,** 368-372.

Hoffman, W., Lakkis, F. G. & Chalasani, G. 2016. B Cells, Antibodies, and More. *Clinical Journal of the American Society of Nephrology,* 11**,** 137-154.

Honegger, A. & Plückthun, A. 2001. Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. *Journal of Molecular Biology,* 309**,** 657-70.

Horns, F., Vollmers, C., Croote, D., Mackey, S. F., Swan, G. E., Dekker, C. L., Davis, M. M. & Quake, S. R. 2016. Lineage tracing of human B cells reveals the in vivo landscape of human antibody class switching. *eLife,* 5.

Howell, K. A., Qiu, X., Brannan, J. M., Bryan, C., Davidson, E., Holtsberg, F. W., Wec, A. Z., Shulenin, S., Biggins, J. E., Douglas, R., Enterlein, S. G., Turner, H. L., Pallesen, J., Murin, C. D., He, S., Kroeker, A., Vu, H., Herbert, A. S., Fusco, M. L., Nyakatura, E. K., Lai, J. R., Keck, Z. Y., Foung, S. K. H., Saphire, E. O., Zeitlin, L., Ward, A. B., Chandran, K., Doranz, B. J., Kobinger, G. P., Dye, J. M. & Aman, M. J. 2016. Antibody Treatment of Ebola and Sudan Virus Infection via a Uniquely Exposed Epitope within the Glycoprotein Receptor-Binding Site. *Cell Reports,* 15**,** 1514-1526.

IJspeert, H., van Schouwenburg, P. A., van Zessen, D., Pico-Knijnenburg, I., Stubbs, A. P. & van der Burg, M. 2017. Antigen Receptor Galaxy: A User-Friendly, Web-Based Tool for Analysis and Visualization of T and B Cell Receptor Repertoire Data. *Journal of Immunology,* 198**,** 4156-4165.

Ikai, A. 1980. Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry,* 88**,** 1895-1898.

Irani, V., Guy, A. J., Andrew, D., Beeson, J. G., Ramsland, P. A. & Richards, J. S. 2015. Molecular properties of human IgG subclasses and their implications for designing therapeutic monoclonal antibodies against infectious diseases. *Molecular Immunology,* 67**,** 171-182.

Ito, H., Watanabe, S., Takada, A. & Kawaoka, Y. 2001. Ebola Virus Glycoprotein: Proteolytic Processing, Acylation, Cell Tropism, and Detection of Neutralizing Antibodies. *Journal of Virology,* 75**,** 1576-1580.

Jackson, K. J., Wang, Y. & Collins, A. M. 2014. Human immunoglobulin classes and subclasses show variability in VDJ gene mutation levels. *Immunology and Cell Biology,* 92**,** 729-33.

James, L. C., Roversi, P. & Tawfik, D. S. 2003. Antibody multispecificity mediated by conformational diversity. *Science,* 299**,** 1362-1367.

James, L. C. & Tawfik, D. S. 2003. The specificity of cross-reactivity: Promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Science,* 12**,** 2183-2193.

James, L. K., Shamji, M. H., Walker, S. M., Wilson, D. R., Wachholz, P. A., Francis, J. N., Jacobson, M. R., Kimber, I., Till, S. J. & Durham, S. R. 2011. Long-term tolerance after allergen immunotherapy is accompanied by selective persistence of blocking antibodies. *Journal of Allergy and Clinical Immunology,* 127**,** 509-516.e1-5.

Janda, A., Bowen, A., Greenspan, N. S. & Casadevall, A. 2016. Ig Constant Region Effects on Variable Region Structure and Function. *Frontiers in Microbiology,* 7**,** 22.

Janezic, A., Chapman, C. J., Snow, R. E., Hourihane, J. O., Warner, J. O. & Stevenson, F. K. 1998. Immunogenetic analysis of the heavy chain variable regions of IgE from patients allergic to peanuts. *Journal of Allergy and Clinical Immunology,* 101**,** 391-6.

Jayaram, N., Bhowmick, P. & Martin, A. C. R. 2012. Germline V_H/V_L pairing in antibodies. *Protein Engineering Design & Selection,* 25**,** 523-529.

Jefferis, R. 2012. Isotype and glycoform selection for antibody therapeutics. *Archives of Biochemistry and Biophysics,* 526**,** 159-66.

Jeffers, S. A., Sanders, D. A. & Sanchez, A. 2002. Covalent modifications of the ebola virus glycoprotein. *Journal of Virology,* 76**,** 12463-72.

Jørgensen, C. S., Hansen, K. B., Jacobssen, S., Halberg, P., Ullman, S., Hansen, D., Mikkelsen, T. L., Weile, B., Madsen, M. H., Wiik, A. & Houen, G. 2005. Absence of high-affinity calreticulin autoantibodies in patients with systemic rheumatic diseases and coeliac disease. *Scandinavian Journal of Clinical & Laboratory Investigation,* 65**,** 403-412.

Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. 1983. *Sequence of Proteins of Immunological Interest,* Bethesda, National Institutes of Health.

Kabsch, W. & Sander, C. 1983. Dictionary of protein secondary structure: Pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers,* 22**,** 2577-2637.

Kalinina, O., Doyle-Cooper, C. M., Miksanek, J., Meng, W., Prak, E. L. & Weigert, M. G. 2011. Alternative mechanisms of receptor editing in autoreactive B cells. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 7125-7130.

Kalinina, O., Wang, Y., Sia, K., Radic, M., Cazenave, P. A. & Weigert, M. 2014. Light chain editors of anti-DNA receptors in human B cells. *Journal of Experimental Medicine,* 211**,** 357-364.

Keck, Z. Y., Enterlein, S. G., Howell, K. A., Vu, H., Shulenin, S., Warfield, K. L., Froude, J. W., Araghi, N., Douglas, R., Biggins, J., Lear-Rooney, C. M., Wirchnianski, A. S., Lau, P., Wang, Y., Herbert, A. S., Dye, J. M., Glass, P. J., Holtsberg, F. W., Foung, S. K. & Aman, M. J. 2016. Macaque Monoclonal Antibodies Targeting Novel Conserved Epitopes within Filovirus Glycoprotein. *Journal of Virology,* 90**,** 279-91.

Keitel, T., Kramer, A., Wessner, H., Scholz, C., Schneider-Mergener, J. & Hohne, W. 1997. Crystallographic analysis of anti-p24 (HIV-1) monoclonal antibody cross-reactivity and polyspecificity. *Cell,* 91**,** 811-20.

Khurana, S., Fuentes, S., Coyle, E. M., Ravichandran, S., Davey, R. T., Jr. & Beigel, J. H. 2016. Human antibody repertoire after VSV-Ebola vaccination identifies novel targets and virus-neutralizing IgM antibodies. *Nature Medicine,* 22**,** 1439-1447.

Kidera, A., Konishi, Y., Oka, M., Ooi, T. & Scheraga, H. A. 1985a. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *Journal of Protein Chemistry,* 4**,** 23-55.

Kidera, A., Konishi, Y., Ooi, T. & Scheraga, H. A. 1985b. Relation between sequence similarity and structural similarity in proteins - role of important properties of amino acids. *Journal of Protein Chemistry,* 4**,** 265-297.

Kitaura, K., Yamashita, H., Ayabe, H., Shini, T., Matsutani, T. & Suzuki, R. 2017. Different Somatic Hypermutation Levels among Antibody Subclasses Disclosed by a New Next-Generation Sequencing-Based Antibody Repertoire Analysis. *Frontiers in Immunology,* 8**,** 389.

Köhler, F., Hug, E., Eschbach, C., Meixlsperger, S., Hobeika, E., Kofer, J., Wardemann, H. & Jumaa, H. 2008. Autoreactive B cell receptors mimic autonomous pre-B cell receptor signaling and induce proliferation of early B cells. *Immunity,* 29**,** 912-21.

Kovaltsuk, A., Krawczyk, K., Galson, J. D., Kelly, D. F., Deane, C. M. & Trück, J. 2017. How B-Cell Receptor Repertoire Sequencing Can Be Enriched with Structural Antibody Data. *Frontiers in Immunology,* 8**,** 1753.

Kramer, A., Keitel, T., Winkler, K., Stöcklein, W., Höhne, W. & Schneider-Mergener, J. 1997. Molecular basis for the binding promiscuity of an anti-p24 (HIV-1) monoclonal antibody. *Cell,* 91**,** 799-809.

Kreuels, B., Wichmann, D., Emmerich, P., Schmidt-Chanasit, J., de Heer, G., Kluge, S., Sow, A., Renné, T., Günther, S., Lohse, A. W., Addo, M. M. & Schmiedel, S. 2014. A Case of Severe Ebola Virus Infection Complicated by Gram-Negative Septicemia. *New England Journal of Medicine,* 371**,** 2394-2401.

Ksiazek, T. G., Rollin, P. E., Williams, A. J., Bressler, D. S., Martin, M. L., Swanepoel, R., Burt, F. J., Leman, P. A., Khan, A. S., Rowe, A. K., Mukunu, R., Sanchez, A. & Peters, C. J. 1999. Clinical virology of Ebola hemorrhagic fever (EHF): virus, virus antigen, and IgG and IgM antibody findings among EHF patients in Kikwit, Democratic Republic of the Congo, 1995. *Journal of Infectious Diseases,* 179 Suppl 1**,** S177-87.

Kunik, V. & Ofran, Y. 2013. The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops. *Protein Engineering, Design & Selection,* 26**,** 599-609.

Kyte, J. & Doolittle, R. F. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology,* 157**,** 105-132.

Laffy, J. M. J., Dodev, T., Macpherson, J. A., Townsend, C., Lu, H. C., Dunn-Walters, D. & Fraternali, F. 2016. Promiscuous antibodies characterised by their physico-chemical properties: From sequence to structure and back. *Progress in Biophysics and Molecular Biology*.

Larimore, K., McCormick, M. W., Robins, H. S. & Greenberg, P. D. 2012. Shaping of Human Germline IgH Repertoires Revealed by Deep Sequencing. *Journal of Immunology,* 189**,** 3221-3230.

Lavinder, J. J., Horton, A. P., Georgiou, G. & Ippolito, G. C. 2015. Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. *Current Opinion in Chemical Biology,* 24**,** 112-120.

Lee, C. M., Iorno, N., Sierro, F. & Christ, D. 2007. Selection of human antibody fragments by phage display. *Nature Protocols,* 2**,** 3001-8.

Lee, J. E., Fusco, M. L., Hessell, A. J., Oswald, W. B., Burton, D. R. & Saphire, E. O. 2008. Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature,* 454**,** 177-82.

Lee, J. E. & Saphire, E. O. 2009. Neutralizing ebolavirus: structural insights into the envelope glycoprotein and antibodies targeted against it. *Current Opinion in Structural Biology,* 19**,** 408-17.

Lefranc, M.-P., Pommié, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., Thouvenin-Contet, V. & Lefranc, G. 2003. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Developmental & Comparative Immunology,* 27**,** 55-77.

Lefranc, M. P. 1997. Unique database numbering system for immunogenetic analysis. *Immunology Today,* 18**,** 509.

Lefranc, M. P. 2014. Immunoglobulin and T Cell Receptor Genes: IMGT and the Birth and Rise of Immunoinformatics. *Frontiers in Immunology,* 5**,** 22.

Lefranc, M. P. 2016. *IG and TR number of genes: Human* [Online]. Available: http://www.imgt.org/IMGTrepertoire/LocusGenes/genetable/human/geneNumber.html#refs [Accessed 03/06/2018 2018].

Lefranc, M. P. & Lefranc, G. 2001. *The Immunoglobulin FactsBook,* San Diego, CA, Academic Press.

Lennemann, N. J., Rhein, B. A., Ndungo, E., Chandran, K., Qiu, X. & Maury, W. 2014. Comprehensive functional analysis of N-linked glycans on Ebola virus GP1. *mBio,* 5**,** e00862-13.

Leroy, E. M., Baize, S., Debre, P., Lansoud-Soukate, J. & Mavoungou, E. 2001. Early immune responses accompanying human asymptomatic Ebola infections. *Clinical and Experimental Immunology,* 124**,** 453-460.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Physics,* 10**,** 707–710.

Li, H., Jiang, Y. F., Prak, E. L., Radic, M. & Weigert, M. 2001. Editors and editing of anti-DNA receptors. *Immunity,* 15**,** 947-957.

Li, S., Lefranc, M. P., Miles, J. J., Alamyar, E., Giudicelli, V., Duroux, P., Freeman, J. D., Corbin, V. D., Scheerlinck, J. P., Frohman, M. A., Cameron, P. U., Plebanski, M., Loveland, B., Burrows, S. R., Papenfuss, A. T. & Gowans, E. J. 2013. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nature Communications,* 4**,** 2333.

Liang, Z., Xie, C., Chen, C., Kreska, D., Hsu, K., Li, L., Zhou, X. J. & Mohan, C. 2004. Pathogenic Profiles and Molecular Signatures of Antinuclear Autoantibodies Rescued from NZM2410 Lupus Mice. *The Journal of Experimental Medicine,* 199**,** 381-398.

Liu, H. & May, K. 2012. Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs,* 4**,** 17-23.

Lydyard, P., Whelan, A. & Fanger, M. 2011. *BIOS Instant Notes in Immunology*, Garland Science, Taylor & Francis Group, LLC.

Ma, B., Shatsky, M., Wolfson, H. J. & Nussinov, R. 2002. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Science,* 11**,** 184-97.

MacNeil, A., Reed, Z. & Rollin, P. E. 2011. Serologic cross-reactivity of human IgM and IgG antibodies to five species of Ebola virus. *PLoS Neglected Tropical Diseases,* 5**,** e1175.

Malcolm, S., Barton, P., Murphy, C., Fergusonsmith, M. A., Bentley, D. L. & Rabbitts, T. H. 1982. Localization of human-immunoglobulin kappa-light chain variable region genes to the short arm of chromosome-2 by insitu hybridization. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 79**,** 4957-4961.

Mamedov, I. Z., Britanova, O. V., Zvyagin, I. V., Turchaninova, M. A., Bolotin, D. A., Putintseva, E. V., Lebedev, Y. B. & Chudakov, D. M. 2013. Preparing Unbiased T-Cell Receptor and Antibody cDNA Libraries for the Deep Next Generation Sequencing Profiling. *Frontiers in Immunology,* 4**,** 456.

Margo, C. E. & Harman, L. E. 2016. Autoimmune disease: Conceptual history and contributions of ocular immunology. *Survey of Ophthalmology,* 61**,** 680-8.

Margreitter, C., Lu, H. C., Townsend, C., Stewart, A., Dunn-Walters, D. K. & Fraternali, F. 2018. BRepertoire: a user-friendly web server for analysing antibody repertoire data. *Nucleic Acids Research,* 46**,** W264-W270.

Martin, V. G., Wu, Y. C. B., Townsend, C. L., Lu, G. H., Silva O'Hare, J., Mozeika, A., Coolen, A. C., Kipling, D., Fraternali, F. & Dunn-Walters, D. K. 2016. Transitional B cells in early human B cell development - time to revisit the paradigm? *Frontiers in Immunology,* 7**,** 13.

Maruyama, T., Rodriguez, L. L., Jahrling, P. B., Sanchez, A., Khan, A. S., Nichol, S. T., Peters, C. J., Parren, P. W. & Burton, D. R. 1999. Ebola virus can be effectively neutralized by antibody produced in natural human infection. *Journal of Virology,* 73**,** 6024-30.

Maul, R. W. & Gearhart, P. J. 2010. AID and Somatic Hypermutation. *Advances in Immunology,* 105**,** 159-191.

McDaniel, J. R., DeKosky, B. J., Tanno, H., Ellington, A. D. & Georgiou, G. 2016. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nature Protocols,* 11**,** 429-442.

McElroy, A. K., Akondy, R. S., Davis, C. W., Ellebedy, A. H., Mehta, A. K., Kraft, C. S., Lyon, G. M., Ribner, B. S., Varkey, J., Sidney, J., Sette, A., Campbell, S., Ströher, U., Damon, I., Nichol, S. T., Spiropoulou, C. F. & Ahmed, R. 2015. Human Ebola virus infection results in substantial immune activation. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** 4719-24.

Meffre, E., Schaefer, A., Wardemann, H., Wilson, P., Davis, E. & Nussenzweig, M. C. 2004. Surrogate light chain expressing human peripheral B cells produce self-reactive antibodies. *Journal of Experimental Medicine,* 199**,** 145-150.

Mehedi, M., Falzarano, D., Seebach, J., Hu, X., Carpenter, M. S., Schnittler, H.-J. & Feldmann, H. 2011. A New Ebola Virus Nonstructural Glycoprotein Expressed through RNA Editing. *Journal of Virology,* 85**,** 5406-5414.

Meyers, G., Ng, Y. S., Bannock, J. M., Lavoie, A., Walter, J. E., Notarangelo, L. D., Kilic, S. S., Aksu, G., Debré, M., Rieux-Laucat, F., Conley, M. E., Cunningham-Rundles, C., Durandy, A. & Meffre, E. 2011. Activation-induced cytidine deaminase (AID) is required for B-cell tolerance in humans. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 11554-9.

Misasi, J., Gilman, M. S., Kanekiyo, M., Gui, M., Cagigi, A., Mulangu, S., Corti, D., Ledgerwood, J. E., Lanzavecchia, A., Cunningham, J., Muyembe-Tamfun, J. J., Baxa, U., Graham, B. S., Xiang, Y., Sullivan, N. J. & McLellan, J. S. 2016. Structural and molecular basis for Ebola virus neutralization by protective human antibodies. *Science,* 351**,** 1343-6.

Mole, C. M., Bene, M. C., Montagne, P. M., Seilles, E. & Faure, G. C. 1994. Light-chains of immunoglobulins in human secretions. *Clinica Chimica Acta,* 224**,** 191-197.

Montalbano, A., Ogwaro, K. M., Tang, A., Matthews, A. G., Larijani, M., Oettinger, M. A. & Feeney, A. J. 2003. V(D)J recombination frequencies can be profoundly affected by changes in the spacer sequence. *Journal of Immunology,* 171**,** 5296-304.

Montano, R. F. & Morrison, S. L. 2002. Influence of the isotype of the light chain on the properties of IgG. *Journal of Immunology,* 168**,** 224-231.

Motea, E. A. & Berdis, A. J. 2010. Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. *Biochimica et Biophysica Acta-Proteins and Proteomics,* 1804**,** 1151-1166.

Mouquet, H., Scheid, J. F., Zoller, M. J., Krogsgaard, M., Ott, R. G., Shukair, S., Artyomov, M. N., Pietzsch, J., Connors, M., Pereyra, F., Walker, B. D., Ho, D. D., Wilson, P. C., Seaman, M. S., Eisen, H. N., Chakraborty, A. K., Hope, T. J., Ravetch, J. V., Wardemann, H. & Nussenzweig, M. C. 2010. Polyreactivity increases the apparent affinity of anti-HIV antibodies by heteroligation. *Nature,* 467**,** 591-5.

Munshaw, S. & Kepler, T. B. 2010. SoDA2: a Hidden Markov Model approach for identification of immunoglobulin rearrangements. *Bioinformatics,* 26**,** 867-72.

Murin, C. D., Fusco, M. L., Bornholdt, Z. A., Qiu, X., Olinger, G. G., Zeitlin, L., Kobinger, G. P., Ward, A. B. & Saphire, E. O. 2014. Structures of protective antibodies reveal sites of vulnerability on Ebola virus. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 17182-17187.

Nadel, B., Tang, A., Escuro, G., Lugo, G. & Feeney, A. J. 1998. Sequence of the Spacer in the Recombination Signal Sequence Affects V(D)J Rearrangement Frequency and Correlates with Nonrandom Vκ Usage In Vivo. *The Journal of Experimental Medicine,* 187**,** 1495-1503.

Nemazee, D. 2006. Receptor editing in lymphocyte development and central tolerance. *Nature Reviews Immunology,* 6**,** 728-740.

Nemazee, D. 2017. Mechanisms of central tolerance for B cells. *Nature Reviews Immunology,* 17**,** 281-294.

Nemazee, D. A. & Bürki, K. 1989. Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. *Nature,* 337**,** 562-6.

Ng, Y.-S., Wardemann, H., Chelnis, J., Cunningham-Rundles, C. & Meffre, E. 2004. Bruton's Tyrosine Kinase Is Essential for Human B Cell Tolerance. *The Journal of Experimental Medicine,* 200**,** 927-934.

Nguyen, V. K., Hamers, R., Wyns, L. & Muyldermans, S. 2000. Camel heavy-chain antibodies: diverse germline VHH and specific mechanisms enlarge the antigen-binding repertoire. *EMBO Journal,* 19**,** 921-930.

Nirula, A., Glaser, S. M., Kalled, S. L. & Taylor, F. R. 2011. What is IgG4? A review of the biology of a unique immunoglobulin subtype. *Current Opinion in Rheumatology,* 23**,** 119-24.

North, B., Lehmann, A. & Dunbrack, R. L., Jr. 2011. A new clustering of antibody CDR loop conformations. *Journal of Molecular Biology,* 406**,** 228-56.

Notkins, A. L. 2004. Polyreactivity of antibody molecules. *Trends in Immunology,* 25**,** 174-179.

Nouri-Aria, K. T., Wachholz, P. A., Francis, J. N., Jacobson, M. R., Walker, S. M., Wilcock, L. K., Staple, S. Q., Aalberse, R. C., Till, S. J. & Durham, S. R. 2004. Grass pollen immunotherapy induces mucosal and peripheral IL-10 responses and blocking IgG activity. *Journal of Immunology,* 172**,** 3252-9.

Obukhanych, T. V. & Nussenzweig, M. C. 2006. T-independent type II immune responses generate memory B cells. *The Journal of Experimental Medicine,* 203**,** 305-310.

Ohm-Laursen, L., Nielsen, M., Larsen, S. R. & Barington, T. 2006. No evidence for the use of DIR, D–D fusions, chromosome 15 open reading frames or V(H)replacement in the peripheral repertoire was found on application of an improved algorithm, JointML, to 6329 human immunoglobulin H rearrangements. *Immunology,* 119**,** 265-277.

Oldstone, M. B. 1998. Molecular mimicry and immune-mediated diseases. *FASEB Journal,* 12**,** 1255-65.

Osorio, D., Rondon-Villarreal, P. & Torres, R. 2015. Peptides: Calculate Indices and Theoretical Properties of Protein Sequences. R package version 1.1.1.

Oswald, W. B., Geisbert, T. W., Davis, K. J., Geisbert, J. B., Sullivan, N. J., Jahrling, P. B., Parren, P. W. & Burton, D. R. 2007. Neutralizing antibody fails to impact the course of Ebola virus infection in monkeys. *PLoS Pathogens,* 3**,** e9.

Owen, J., Punt, J. & Stranford, S. 2013. *Kuby Immunology,* New York, W.H. Freeman and Company.

Padlan, E. A. 1994. Anatomy of the antibody molecule. *Molecular Immunology,* 31**,** 169-217.

Palanichamy, A., Barnard, J., Zheng, B., Owen, T., Quach, T., Wei, C. W., Looney, R. J., Sanz, I. & Anolik, J. H. 2009. Novel Human Transitional B Cell Populations Revealed by B Cell Depletion Therapy. *Journal of Immunology,* 182**,** 5982-5993.

Pancera, M., McLellan, J. S., Wu, X., Zhu, J., Changela, A., Schmidt, S. D., Yang, Y., Zhou, T., Phogat, S., Mascola, J. R. & Kwong, P. D. 2010. Crystal structure of PG16 and chimeric dissection with somatically related PG9: structure-function analysis of two quaternary-specific antibodies that effectively neutralize HIV-1. *Journal of Virology,* 84**,** 8098-110.

Parham, P. 2009. *The Immune System,* New York, Garland Science.

Parren, P. W. H. I., Geisbert, T. W., Maruyama, T., Jahrling, P. B. & Burton, D. R. 2002. Pre- and Postexposure Prophylaxis of Ebola Virus Infection in an Animal Model by Passive Transfer of a Neutralizing Human Antibody. *Journal of Virology,* 76**,** 6408-6412.

Persson, H., Ye, W., Wernimont, A., Adams, J. J., Koide, A., Koide, S., Lam, R. & Sidhu, S. S. 2013. CDR-H3 Diversity Is Not Required for Antigen Recognition by Synthetic Antibodies. *Journal of Molecular Biology,* 425**,** 803-811.

Potter, K. N., Hobby, P., Klijn, S., Stevenson, F. K. & Sutton, B. J. 2002. Evidence for involvement of a hydrophobic patch in framework region 1 of human V4-34-encoded Igs in recognition of the red blood cell I antigen. *Journal of Immunology,* 169**,** 3777-82.

Qiu, X., Alimonti, J. B., Melito, P. L., Fernando, L., Ströher, U. & Jones, S. M. 2011. Characterization of Zaire ebolavirus glycoprotein-specific monoclonal antibodies. *Clinical Immunology,* 141**,** 218-27.

Qiu, X., Audet, J., Wong, G., Pillet, S., Bello, A., Cabral, T., Strong, J. E., Plummer, F., Corbett, C. R., Alimonti, J. B. & Kobinger, G. P. 2012. Successful Treatment of Ebola Virus-Infected Cynomolgus Macaques with Monoclonal Antibodies. *Science Translational Medicine,* 4**,** 138ra81.

Qiu, X., Wong, G., Audet, J., Bello, A., Fernando, L., Alimonti, J. B., Fausther-Bovendo, H., Wei, H., Aviles, J., Hiatt, E., Johnson, A., Morton, J., Swope, K., Bohorov, O., Bohorova, N., Goodman, C., Kim, D., Pauly, M. H., Velasco, J., Pettitt, J., Olinger, G. G., Whaley, K., Xu, B., Strong, J. E., Zeitlin, L. & Kobinger, G. P. 2014. Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. *Nature,* 514**,** 47-53.

Raaphorst, F. M., Raman, C. S., Tami, J., Fischbach, M. & Sanz, I. 1997. Human Ig heavy chain CDR3 regions in adult bone marrow pre-B cells display an adult phenotype of diversity: evidence for structural selection of D-H amino acid sequences. *International Immunology,* 9**,** 1503-1515.

Rackovsky, S. & Scheraga, H. A. 2011. On the Information Content of Protein Sequences. *Journal of Biomolecular Structure & Dynamics,* 28**,** 593-594.

Rajan, S., Kierny, M. R., Mercer, A., Wu, J., Tovchigrechko, A., Wu, H., Dall'Acqua, W. F., Xiao, X. & Chowdhury, P. S. 2018. Recombinant human B cell repertoires enable screening for rare, specific, and natively paired antibodies. *Communications Biology,* 1**,** 5.

Ramsden, D. A. & Wu, G. E. 1991. Mouse kappa light-chain recombination signal sequences mediate recombination more frequently than do those of lambda light chain. *Proceedings of the National Academy of Sciences of the United States of America,* 88**,** 10721-10725.

Rasband, W. S. 1997-2016. ImageJ. Bethesda, Maryland, USA: U. S. National Institutes of Health.

Raux, M., Finkielsztejn, L., Salmon-Céron, D., Bouchez, H., Excler, J. L., Dulioust, E., Grouin, J. M., Sicard, D. & Blondeau, C. 2000. IgG Subclass Distribution in Serum and Various Mucosal Fluids of HIV Type 1-Infected Subjects. *AIDS Research and Human Retroviruses,* 16**,** 583-594.

Retter, I., Althaus, H. H., Münch, R. & Müller, W. 2005. VBASE2, an integrative V gene database. *Nucleic Acids Research,* 33**,** D671-D674.

Rhoads, A. & Au, K. F. 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics,* 13**,** 278-289.

Rice, P., Longden, I. & Bleasby, A. 2000. EMBOSS: The European molecular biology open software suite. *Trends in Genetics,* 16**,** 276-277.

Rini, J. M., Schulze-Gahmen, U. & Wilson, I. A. 1992. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science,* 255**,** 959-65.

Rispens, T., Ooijevaar-de Heer, P., Bende, O. & Aalberse, R. C. 2011. Mechanism of immunoglobulin G4 Fab-arm exchange. *Journal of the American Chemical Society,* 133**,** 10302-11.

Robin, G., Sato, Y., Desplancq, D., Rochel, N., Weiss, E. & Martineau, P. 2014. Restricted diversity of antigen binding residues of antibodies revealed by computational alanine scanning of 227 antibody-antigen complexes. *Journal of Molecular Biology,* 426**,** 3729-3743.

Roos, A., Bouwman, L. H., van Gijlswijk-Janssen, D. J., Faber-Krol, M. C., Stahl, G. L. & Daha, M. R. 2001. Human IgA activates the complement system via the mannan-binding lectin pathway. *Journal of Immunology,* 167**,** 2861-8.

Sack, U., Conrad, K., Csernok, E., Frank, I., Hiepe, F., Krieger, T., Kromminga, A., von Landenberg, P., Messer, G., Witte, T. & Mierau, R. 2009. Autoantibody detection using indirect immunofluorescence on HEp-2 cells. *Annals of the New York Academy of Sciences,* 1173**,** 166-73.

Sajadi, M. M., Farshidpour, M., Brown, E. P., Ouyang, X., Seaman, M. S., Pazgier, M., Ackerman, M. E., Robinson, H., Tomaras, G., Parsons, M. S., Charurat, M., DeVico, A. L., Redfield, R. R. & Lewis, G. K. 2016. λ Light Chain Bias Associated With Enhanced Binding and Function of Anti-HIV Env Glycoprotein Antibodies. *Journal of Infectious Diseases,* 213**,** 156-164.

Samuels, J., Ng, Y. S., Coupillaud, C., Paget, D. & Meffre, E. 2005. Impaired early B cell tolerance in patients with rheumatoid arthritis. *Journal of Experimental Medicine,* 201**,** 1659-1667.

Sanchez, A., Trappier, S. G., Mahy, B. W., Peters, C. J. & Nichol, S. T. 1996. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 3602-3607.

Sblattero, D. & Bradbury, A. 1998. A definitive set of oligonucleotide primers for amplifying human V regions. *Immunotechnology,* 3**,** 271-8.

Scaviner, D., Barbie, V., Ruiz, M. & Lefranc, M. P. 1999. Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. *Experimental and Clinical Immunogenetics,* 16**,** 234-240.

Schatz, D. G. & Ji, Y. H. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nature Reviews Immunology,* 11**,** 251-263.

Schroeder, H. W. & Cavacini, L. 2010. Structure and Function of Immunoglobulins. *The Journal of Allergy and Clinical Immunology,* 125**,** S41-S52.

Sigounas, G., Harindranath, N., Donadel, G. & Notkins, A. L. 1994. Half-Life of Polyreactive Antibodies. *Journal of Clinical Immunology,* 14**,** 134-140.

Snow, R. E., Chapman, C. J., Frew, A. J., Holgate, S. T. & Stevenson, F. K. 1995. Analysis of Ig VH region genes encoding IgE antibodies in splenic B lymphocytes of a patient with asthma. *Journal of Immunology,* 154**,** 5576-81.

Souto-Carneiro, M. M., Longo, N. S., Russ, D. E., Sun, H. W. & Lipsky, P. E. 2004. Characterization of the human Ig heavy chain antigen binding complementarity determining region 3 using a newly developed software algorithm, JOINSOLVER. *Journal of Immunology,* 172**,** 6790-802.

Stanfield, R. L., Zemla, A., Wilson, I. A. & Rupp, B. 2006. Antibody elbow angles are influenced by their light chain class. *Journal of Molecular Biology,* 357**,** 1566-1574.

Stansell, E., Panico, M., Canis, K., Pang, P.-C., Bouché, L., Binet, D., O'Connor, M.-J., Chertova, E., Bess, J., Lifson, J. D., Haslam, S. M., Morris, H. R., Desrosiers, R. C. & Dell, A. 2015. Gp120 on HIV-1 Virions Lacks O-Linked Carbohydrate. *PLoS ONE,* 10**,** e0124784.

Stavnezer, J., Guikema, J. E. J. & Schrader, C. E. 2008. Mechanism and Regulation of Class Switch Recombination. *Annual Review of Immunology,* 26**,** 261-292.

Stavnezer, J. & Schrader, C. E. 2014. Ig heavy chain class switch recombination: mechanism and regulation. *Journal of Immunology,* 193**,** 5370-5378.

Steinsbø, Ø., Dunand, C. J. H., Huang, M., Mesin, L., Salgado-Ferrer, M., Lundin, K. E. A., Jahnsen, J., Wilson, P. C. & Sollid, L. M. 2014. Restricted VH/VL usage and limited mutations in gluten-specific IgA of coeliac disease lesion plasma cells. *Nature Communications,* 5**,** 4041.

Steinwand, M., Droste, P., Frenzel, A., Hust, M., Dübel, S. & Schirrmann, T. 2014. The influence of antibody fragment format on phage display based affinity maturation of IgG. *mAbs,* 6**,** 204-218.

Sterner, E., Flanagan, N. & Gildersleeve, J. C. 2016. Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database. *ACS Chemical Biology,* 11**,** 1773-83.

Stijlemans, B., Conrath, K., Cortez-Retamozo, V., Van Xong, H., Wyns, L., Senter, P., Revets, H., De Baetselier, P., Muyldermans, S. & Magez, S. 2004. Efficient targeting of conserved cryptic epitopes of infectious agents by single domain antibodies. African trypanosomes as paradigm. *Journal of Biological Chemistry,* 279**,** 1256-61.

Tan, J., Pieper, K., Piccoli, L., Abdi, A., Perez, M. F., Geiger, R., Tully, C. M., Jarrossay, D., Maina Ndungu, F., Wambua, J., Bejon, P., Fregni, C. S., Fernandez-Rodriguez, B., Barbieri, S., Bianchi, S., Marsh, K., Thathy, V., Corti, D., Sallusto, F., Bull, P. & Lanzavecchia, A. 2016. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature,* 529**,** 105-109.

Tan, Y. C., Kongpachith, S., Blum, L. K., Ju, C. H., Lahey, L. J., Lu, D. R., Cai, X., Wagner, C. A., Lindstrom, T. M., Sokolove, J. & Robinson, W. H. 2014. Barcode-enabled sequencing of plasmablast antibody repertoires in rheumatoid arthritis. *Arthritis & Rheumatology,* 66**,** 2706-15.

Tang, J., Wang, L., Markiv, A., Jeffs, S. A., Dreja, H., McKnight, A., He, M. & Kang, A. S. 2012. Accessing of recombinant human monoclonal antibodies from patient libraries by eukaryotic ribosome display. *Human Antibodies,* 21**,** 1-11.

Terato, K., Do, C. T., Cutler, D., Waritani, T. & Shionoya, H. 2014. Preventing intense false positive and negative reactions attributed to the principle of ELISA to re-investigate antibody studies in autoimmune diseases. *Journal of Immunological Methods,* 407**,** 15-25.

Thai, T. H. & Kearney, J. F. 2004. Distinct and opposite activities of human terminal deoxynucleotidyltransferase splice variants. *Journal of Immunology,* 173**,** 4009-19.

Thio, M., Kormelink, T. G., Fischer, M. J., Blokhuis, B. R., Nijkamp, F. P. & Redegeld, F. A. 2012. Antigen Binding Characteristics of Immunoglobulin Free Light Chains: Crosslinking by Antigen is Essential to Induce Allergic Inflammation. *Plos One,* 7**,** 7.

Tiegs, S. L., Russell, D. M. & Nemazee, D. 1993. Receptor Editing in Self-Reactive Bone-Marrow B-Cells. *Journal of Experimental Medicine,* 177**,** 1009-1020.

Tiller, T., Meffre, E., Yurasov, S., Tsuiji, M., Nussenzweig, M. C. & Wardemann, H. 2008. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *Journal of Immunological Methods,* 329**,** 112-124.

Tiller, T., Tsuiji, M., Yurasov, S., Velinzon, K., Nussenzweig, M. C. & Wardemann, H. 2007. Autoreactivity in human IgG^+ memory B cells. *Immunity,* 26**,** 205-213.

Tipton, C. M., Fucile, C. F., Darce, J., Chida, A., Ichikawa, T., Gregoretti, I., Schieferl, S., Hom, J., Jenks, S., Feldman, R. J., Mehr, R., Wei, C., Lee, F. E., Cheung, W. C., Rosenberg, A. F. & Sanz, I. 2015. Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. *Nature Immunology,* 16**,** 755-65.

Toellner, K.-M., Jenkinson, W. E., Taylor, D. R., Khan, M., Sze, D. M. Y., Sansom, D. M., Vinuesa, C. G. & MacLennan, I. C. M. 2002. Low-level Hypermutation in T Cell–independent Germinal Centers Compared with High Mutation Rates Associated with T Cell–dependent Germinal Centers. *The Journal of Experimental Medicine,* 195**,** 383-389.

Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature,* 302**,** 575-81.

Towner, J. S., Khristova, M. L., Sealy, T. K., Vincent, M. J., Erickson, B. R., Bawiec, D. A., Hartman, A. L., Comer, J. A., Zaki, S. R., Ströher, U., Gomes da Silva, F., del Castillo, F., Rollin, P. E., Ksiazek, T. G. & Nichol, S. T. 2006. Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. *Journal of Virology,* 80**,** 6497-516.

Townsend, C., Laffy, J., Wu, Y.-C., Silva O'Hare, J., Martin, V., Kipling, D., Fraternali, F. & Dunn-Walters, D. 2016. Significant differences in physicochemical properties of human immunoglobulin kappa and lambda CDR3 regions. *Frontiers in Immunology,* 7**,** 388.

Tsuiji, M., Yurasov, S., Velinzon, K., Thomas, S., Nussenzweig, M. C. & Wardemann, H. 2006. A checkpoint for autoreactivity in human IgM^+ memory B cell development. *Journal of Experimental Medicine,* 203**,** 393-400.

van de Bovenkamp, F. S., Hafkenscheid, L., Rispens, T. & Rombouts, Y. 2016. The Emerging Importance of IgG Fab Glycosylation in Immunity. *Journal of Immunology,* 196**,** 1435-41.

Van der Stoep, N., Van der Linden, J. & Logtenberg, T. 1993. Molecular evolution of the human immunoglobulin E response: high incidence of shared mutations and clonal relatedness

among epsilon VH5 transcripts from three unrelated patients with atopic dermatitis. *The Journal of Experimental Medicine,* 177**,** 99-107.

Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J. & Johnson, K. S. 1996. Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature Biotechnology,* 14**,** 309-314.

Victora, G. D. & Nussenzweig, M. C. 2012. Germinal centers. *Annual Review of Immunology,* 30**,** 429-57.

Vidarsson, G., Dekkers, G. & Rispens, T. 2014. IgG subclasses and allotypes: from structure to effector functions. *Frontiers in Immunology,* 5**,** 520.

Volpe, J. M., Cowell, L. G. & Kepler, T. B. 2006. SoDA: implementation of a 3D alignment algorithm for inference of antigen receptor recombinations. *Bioinformatics,* 22**,** 438-44.

Volpe, J. M. & Kepler, T. B. 2008. Large-scale analysis of human heavy chain V(D)J recombination patterns. *Immunome Research,* 4**,** 3.

Volpe, J. M. & Kepler, T. B. 2009. Genetic correlates of autoreactivity and autoreactive potential in human Ig heavy chains. *Immunome Research,* 5**,** 1.

Vos, Q., Lees, A., Wu, Z. Q., Snapper, C. M. & Mond, J. J. 2000. B-cell activation by T-cell-independent type 2 antigens as an integral part of the humoral immune response to pathogenic microorganisms. *Immunological Reviews,* 176**,** 154-70.

Wang, G. L. & Dunbrack, R. L. 2003. PISCES: a protein sequence culling server. *Bioinformatics,* 19**,** 1589-1591.

Wang, X., Wu, D., Zheng, S., Sun, J., Tao, L., Li, Y. & Cao, Z. 2008. Ab-origin: an enhanced tool to identify the sourcing gene segments in germline for rearranged antibodies. *BMC Bioinformatics,* 9**,** S20-S20.

Wang, Y. H., Zhang, Z., Burrows, P. D., Kubagawa, H., Bridges, S. L., Jr., Findley, H. W. & Cooper, M. D. 2003. V(D)J recombinatorial repertoire diversification during intraclonal pro-B to B-cell differentiation. *Blood,* 101**,** 1030-7.

Wardemann, H., Hammersen, J. & Nussenzweig, M. C. 2004. Human autoantibody silencing by immunoglobulin light chains. *Journal of Experimental Medicine,* 200**,** 191-199.

Wardemann, H., Yurasov, S., Schaefer, A., Young, J. W., Meffre, E. & Nussenzweig, M. C. 2003. Predominant autoantibody production by early human B cell precursors. *Science,* 301**,** 1374-1377.

Weaver, R., Reiling, L., Feng, G., Drew, D. R., Mueller, I., Siba, P. M., Tsuboi, T., Richards, J. S., Fowkes, F. J. I. & Beeson, J. G. 2016. The association between naturally acquired IgG subclass specific antibodies to the PfRH5 invasion complex and protection from Plasmodium falciparum malaria. *Scientific Reports,* 6**,** 33094.

Weber, M. S., Hemmer, B. & Cepok, S. 2011. The role of antibodies in multiple sclerosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* 1812**,** 239-245.

WHO 2016. Ebola Situation Report 10 June 2016. World Health Organisation.

WHO. 2018a. *Ebola virus disease* [Online]. World Health Organization. Available: http://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease [Accessed 08/06/2018].

WHO. 2018b. *Ebola virus disease – Democratic Republic of the Congo* [Online]. World Health Organisation. Available: http://www.who.int/csr/don/17-august-2018-ebola-drc/en/ [Accessed 21/08/2018].

Wilson, J. A., Hevey, M., Bakken, R., Guest, S., Bray, M., Schmaljohn, A. L. & Hart, M. K. 2000. Epitopes involved in antibody-mediated protection from Ebola virus. *Science,* 287**,** 1664-6.

Wilson, P. C., de Bouteiller, O., Liu, Y.-J., Potter, K., Banchereau, J., Capra, J. D. & Pascual, V. 1998. Somatic Hypermutation Introduces Insertions and Deletions into Immunoglobulin V Genes. *The Journal of Experimental Medicine,* 187**,** 59-70.

Woof, J. M. 2013. Immunoglobulin A: Molecular Mechanisms of Function and Role in Immune Defence. *In:* Nimmerjahn, F. (ed.) *Molecular and Cellular Mechanisms of Antibody Activity.* New York: Springer Science+Business Media.

Woof, J. M. & Kerr, M. A. 2004. IgA function – variations on a theme. *Immunology,* 113**,** 175-177.

Wu, Y. C., Kipling, D. & Dunn-Walters, D. 2015. Assessment of B Cell Repertoire in Humans. *Methods in Molecular Biology,* 1343**,** 199-218.

Wu, Y. C., Kipling, D., Leong, H. S., Martin, V., Ademokun, A. A. & Dunn-Walters, D. K. 2010. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood,* 116**,** 1070-1078.

Xiong, H., Dolpady, J., Wabl, M., Curotto de Lafaille, M. A. & Lafaille, J. J. 2012. Sequential class switching is required for the generation of high affinity IgE antibodies. *The Journal of Experimental Medicine,* 209**,** 353-364.

Xu, J. L. & Davis, M. M. 2000. Diversity in the CDR3 region of V_H is sufficient for most antibody specificities. *Immunity,* 13**,** 37-45.

Yachimovich, N., Mostoslavsky, G., Yarkoni, Y., Verbovetski, I. & Eilat, D. 2002. The efficiency of B cell receptor (BCR) editing is dependent on BCR light chain rearrangement status. *European Journal of Immunology,* 32**,** 1164-1174.

Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Research,* 41**,** W34-40.

Yurasov, S., Tiller, T., Tsuiji, M., Velinzon, K., Pascual, V., Wardemann, H. & Nussenzweig, M. C. 2006. Persistent expression of autoantibodies in SLE patients in remission. *Journal of Experimental Medicine,* 203**,** 2255-61.

Yurasov, S., Wardemann, H., Hammersen, J., Tsuiji, M., Meffre, E., Pascual, V. & Nussenzweig, M. C. 2005. Defective B cell tolerance checkpoints in systemic lupus erythematosus. *Journal of Experimental Medicine,* 201**,** 703-711.

Zahnd, C., Amstutz, P. & Plückthun, A. 2007. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. *Nature Methods,* 4**,** 269-79.

Zhang, J. & Yeh, M. 1994. Cloning, sequencing and analyzing of the heavy chain V region genes of human polyreactive antibodies. *Cell Research,* 4**,** 31-46.

Zhang, K., Mills, F. C. & Saxon, A. 1994. Switch circles from IL-4-directed epsilon class switching from human B lymphocytes. Evidence for direct, sequential, and multiple step sequential switch from mu to epsilon Ig heavy chain gene. *Journal of Immunology,* 152**,** 3427-35.

Zhou, Z. H. & Notkins, A. L. 2004. Polyreactive antigen-binding B (PAB(+)) cells are widely distributed and the PAB(+) population consists of both B-1(+) and B-1(-) phenotypes. *Clinical and Experimental Immunology,* 137**,** 88-100.

Zhou, Z. H., Zhang, Y. H., Hu, Y. F., Wahl, L. M., Cisar, J. O. & Notkins, A. L. 2007. The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies. *Cell Host & Microbe,* 1**,** 51-61.

Publications

Publications

1. Dunn-Walters, D., Townsend, C., Sinclair, E., Stewart, A. 2018. Immunoglobulin gene analysis as a tool for investigating human immune responses. *Immunological Reviews*, 284, 132-147.

2. Margreitter, C., Lu, H.C., Townsend, C., Stewart, A., Dunn-Walters, D.K., Fraternali, F. 2018. BRepertoire: a user-friendly web server for analysing antibody repertoire data. *Nucleic Acids Research*, 46, W264-W270.

3. Martin, V.G., Y.C.B. Wu, C.L. Townsend, G.H.C. Lu, J. Silva O'Hare, A. Mozeika, A.C.C. Coolen, D. Kipling, F. Fraternali & D.K. Dunn-Walters. 2016. Transitional B cells in Early Human B cell Development - Time to Revisit the Paradigm? *Frontiers in Immunology*, 7, 546.

4. Townsend, C.L., J.M.J Laffy, Y.C. Wu, J. Silva O'Hare, V. Martin, D. Kipling, F. Fraternali, and D.K. Dunn-Walters. 2016. Significant differences in physicochemical properties of human immunoglobulin kappa and lambda CDR3 regions. *Frontiers in Immunology*, 7, 388.

5. Laffy, J.M.J., T. Dodev, J.A. Macpherson, C. Townsend, H.C. Lu, D. Dunn-Walters and F. Fraternali. 2016. Promiscuous antibodies characterised by their physico-chemical properties: From sequence to structure and back. *Progress in Biophysics and Molecular Biology*, 128, 47-56.